CRYSTAL STRUCTURE DETERMINATION OF TWO NEW RED FLUORESCENT PROTEINS BY X-RAY CRYSTALLOGRAPHY

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI`I AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTERS OF SCIENCE

IN

CHEMISTRY

AUGUST 2017

By

Mian Huang

Thesis Committee:

Ho Leung Ng, Chairperson

Joseph Jarrett

John Head

Keywords: Red fluorescent protein, Crystallization, X-ray diffraction, Chromophore

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Ho Leung Ng, for accepting me into his research group and introducing me into X-ray crystallography study through two projects involving crystal structure determination of two red fluorescent proteins and other side projects. His willingness to share knowledge, enthusiasm on research and patience to every question I ask, no matter how basic it is, greatly encourage me to make a progress in my study. It is my pleasure to learn and work under his guidance. Also, my lab mates in Ng group offers me a lot of technique support in the beginning and the process of the projects. I sincerely appreciate their kind assistance. Finally, I want to thank my parents and my friends for always supporting me to pursue my dream.

ABSTRACT

Recently scientists are interested in developing new variants of red fluorescent proteins (RFPs), because biosensors basing on RFPs show visual superiority in live-cell imaging. In this thesis, the structures of two RFP variants, K78BoF and mRuby0.4-3, were investigated by X-ray crystallography. A monomeric crystal structure of K78BoF was obtained at pH 6 with resolution at 2.13Å, while a dimeric crystal structure of mRuby0.4-3 was determined at pH 8.2 with resolution at 2.63Å. Both of their chromophores are formed by Met-Tyr-Gly with K78BoF in the *cis* state and mRuby0.4-3 in *trans*. Hydrogen bonds connecting water molecules and residues to K78BoF chromophore probably stabilize the conformation of the chromophore and contribute to the red fluorescence emitted. However, the substitution of an unnatural amino acid *p*-boronophenylalanine breaks two hydrogen bonds, causing a deviation of the conformation. Analysis of hydrogen bonds on mRuby0.4-3 chromophore was limited by the relatively low resolution of the crystal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

3D structure	Three-dimensional structure
Å	Ångström
λ_{max}	Maximum emission wavelength
BFP	Blue fluorescent protein
BoF	*p*-boronophenylalanine
CFP	Cyan fluorescent protein
DI water	Deionized water
EBFP	Enhanced blue fluorescent protein
ECFP	Enhanced cyan fluorescent protein
EGFP	Enhanced green fluorescent protein
FP	Fluorescent protein
FRET	Förster resonance energy transfer
GFP	Green fluorescent protein
HOMO	Highest occupied molecular orbital
KO	Kusabira Orange
LUMO	Lowest unoccupied molecular orbital
LVCO	LV cryo oil
MCSG	Midwest Center for Structural Genomics
NMR	Nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction

PDB	Protein Data Bank
RFP	Red fluorescent protein
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
UV	Ultraviolet
XDS	X-ray Detector Software
YFP	Yellow fluorescent protein

BACKGROUND AND INTRODUCTION

1. Naturally existing fluorescent proteins

As a starting point for the history of scientific research on fluorescent proteins (FPs), in 1962, Osamu Shimomura discovered and isolated a green FP (GFP), avGFP from jellyfish *Aequorea Victoria*, as a companion protein to a photoprotein named aequorin [1]. Later he and his research group clarified the relationship between these two proteins, that aequorin can be triggered by Ca^{2+} to emit a blue light with maximum emission wavelength (λ_{max}) at 460nm which avGFP will then absorb at 400 and 480nm followed with reemitting the absorbed energy as green fluorescence with λ_{max} at 509nm [2]. In 1979, he also deduced the structure of the chromophore for avGFP through synthesizing small model molecules and comparing them with the chromophore isolated from a papain digest of heat-denatured avGFP; in this period NMR had not yet been developed for structure determination [3]. His deduction was confirmed later. Although avGFP was successfully crystallized in 1974 [4] and the x-ray diffraction patterns were reported in 1988 [5], the 3D structure was not solved until 1996, reported separately by Ormö et al [6] and Yang et al [7]. avGFP (PDB 1EMA) is composed of an 11-stranded β-barrel threaded by an α -helix running up the axis of the cylinder, and a chromophore is attached to the α-helix and buried almost perfectly in the center of the cylinder (Fig. 1A).

Early research and the crystal structure proved that avGFP chromophore is formed by intramolecular cyclization of Ser65-Tyr66-Gly67 (Fig. 1B) [8]. Then in 1994, Roger Tsien's group expressed avGFP in *E. coli* cells and found that development of fluorescence from avGFP required participation of oxygen without enzymes or cofactors involved, which was confirmed through blocking fresh protein synthesis by puromycin and tetracycline and controlling the admission of air. A hypothesis of the mechanism for the formation of the chromophore was proposed by the group, as shown in Fig. 2 [9]. To improve its fluorescence properties, Tsien and his collaborators developed multiple mutants with brighter fluorescence and different colors including yellow, cyan and blue, which will be discussed with details later here.

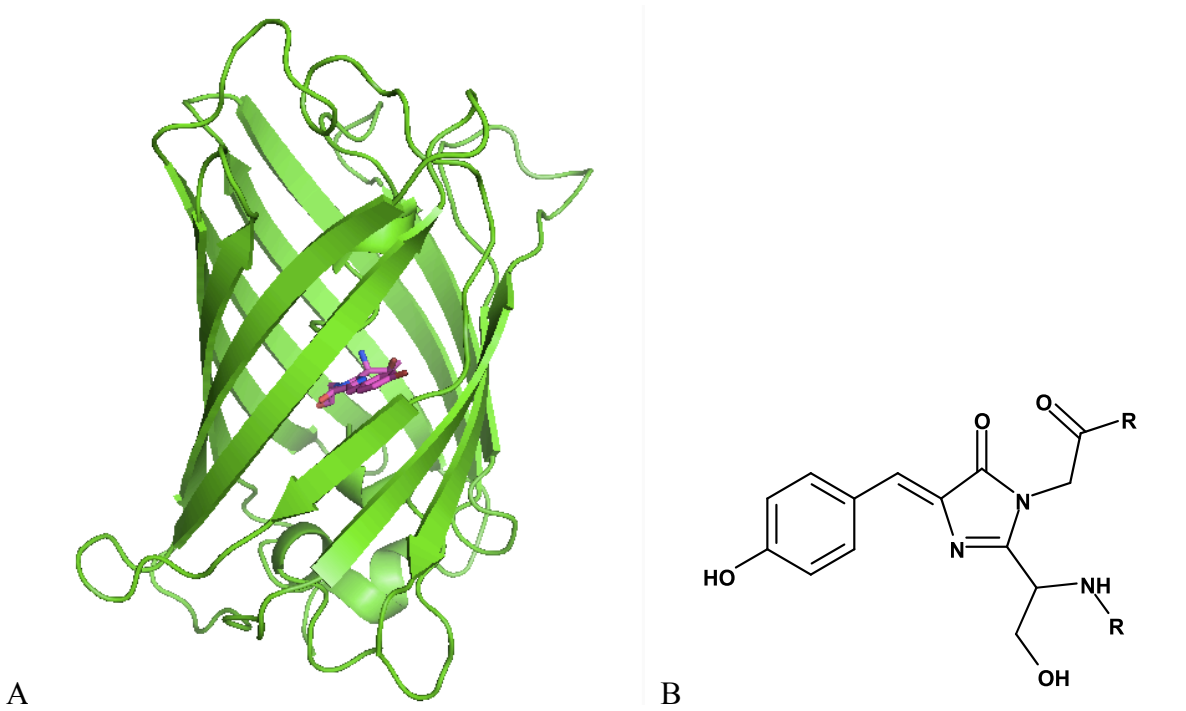


Fig. 1 Structures of green fluorescent protein (PDB 1EMA) (A) and its chromophore (B) isolated from *A. Victoria* [11].

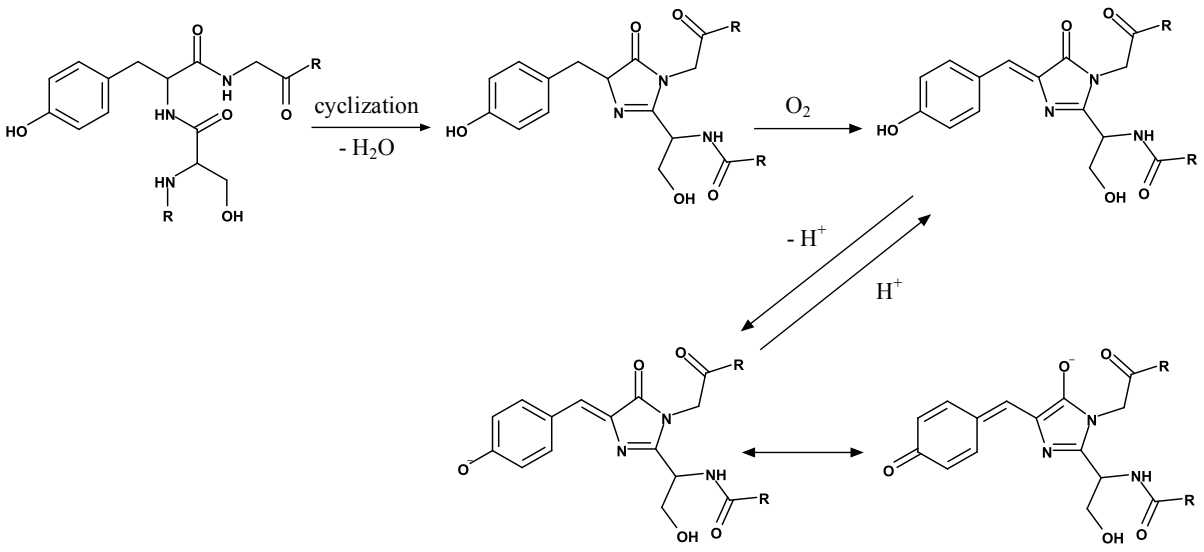


Fig. 2 Proposed biosynthetic scheme for the chromophore of avGFP [9].

Almost at the same time as Tsien's work, Martin Chalfie et al successfully inserted the avGFP gene into *Caenorhabditis elegans* cells and expressed the protein in the first stage *C. elegans* larva without losing the green fluorescence [10]. It showed no cytotoxic effects to the cells in the experiment, which was a milestone for the application of FPs on monitoring gene expression and protein localization in not only prokaryotic but also eukaryotic organisms. In 2008, Osamu Shimomura, Roger Tsien and Martin Chalfie shared the Nobel Prize in Chemistry for their contributions in the discovery and applications of fluorescent proteins.

Wild-type GFP is commonly considered to be the one from *A. Victoria*, but a number of naturally existing GFPs have also been discovered in other species classified in phylum Cnidaria, such as jellyfishes *Mitrocoma*, *Obelia* and *Phialidium* in class Hydrozoa, sea cactus *Cavernularia*, sea pansy *Renilla* and sea pen *Ptilosarcus* and *Pennatula* in class Anthozoa [12]. Later family Pontellidae in phylum Arthropoda [13, 14] and amphioxus in phylum Cephalochordates [15] were also reported to express green FPs. All of them share the same chromophore structure as avGFP with moderate differences in maximum excitation and emission wavelengths, which can be a result from modification of the chemical environment surrounding the chromophore [16-19]. Some GFP-like proteins in these phyla were also found but lack fluorescence [20-22]. For that the species expressing GFP-like proteins located distantly on phylogenetic tree, it raises an interesting research topic to scientists in evolutionary ecology, the diversity and evolution of the GFP family. The production of GFPs, including avGFP, in organisms is generally considered for warning, defense of predator or attraction of prey, while some other functions like photoprotection and protective antioxidation have also been claimed recently. Despite of this, the genes encoding GFPs in different species, especially for the ones coming from different phyla, share little similarity, making it hard to conclude if they originate from a common ancestor or not [23]. With more and more species producing GFPs or GFP-like proteins being discovered and investigated, more clues may be obtained to help answer this question.

FPs emitting other colors have also been discovered in nature. Because they all share the 11-stranded β-barrel structure common to GFPs, all of them belong to the family of GFP-like proteins. Including GFPs, FPs are usually separated into four groups, according to the color difference in human eyes: purple-blue, green, yellow, and orange-red.

So far, over 23 species of GFPs has been isolated that share the same chromophore structure. Despite that most emit green fluorescence upon excitation with λ_{max} at 500-520nm, the subtle difference of the chemical environment surrounding the chromophore affects the excitation and emission wavelengths of the proteins, causing obvious deviation of the color emitted [16]. For example, dsFP483 from *Discosoma striata* is green in the ground state, yet it turns into cyan with λ_{max} at 484nm when it is excited. According to its 2.1Å crystal structure, it shares the same *cis*-coplanar chromophore with avGFP, while the phenolic hydroxyl of its chromophore interacts with Ser146, a water molecule and the imidazole ring of His163 by hydrogen bonds, probably allowing a self-protonation equilibrium that results in blue-shifting [24]. The other cyan FP (CFP) amFP486 from *Anemonia majana* blue-shifting to 486nm could be a consequence of localizing chromophore charge density on the phenolate moiety through interactions between the chromophore and side chains of Ala165 and His199 [25].

Unlike GFPs that are found in diverse species, all other three groups of FPs are only found in limited species of Hydrozoa and Anthozoa. Until now, only one naturally existing yellow FP (YFP) has been reported. zFP538 is a tetramer YFP isolated from *Zoanthus* sp. with a unique three-ring chromophore (Fig. 3a). After Tyr67 and Gly68 form a two-ring structure similar to the GFP chromophore, the side chain of Lys66 is cyclized into a new six-member ring with a cleavage of the peptide backbone between position 65 and 66. An extra C=C bond extends the π-conjugated system and shifts λ_{max} to 539nm [26, 27].

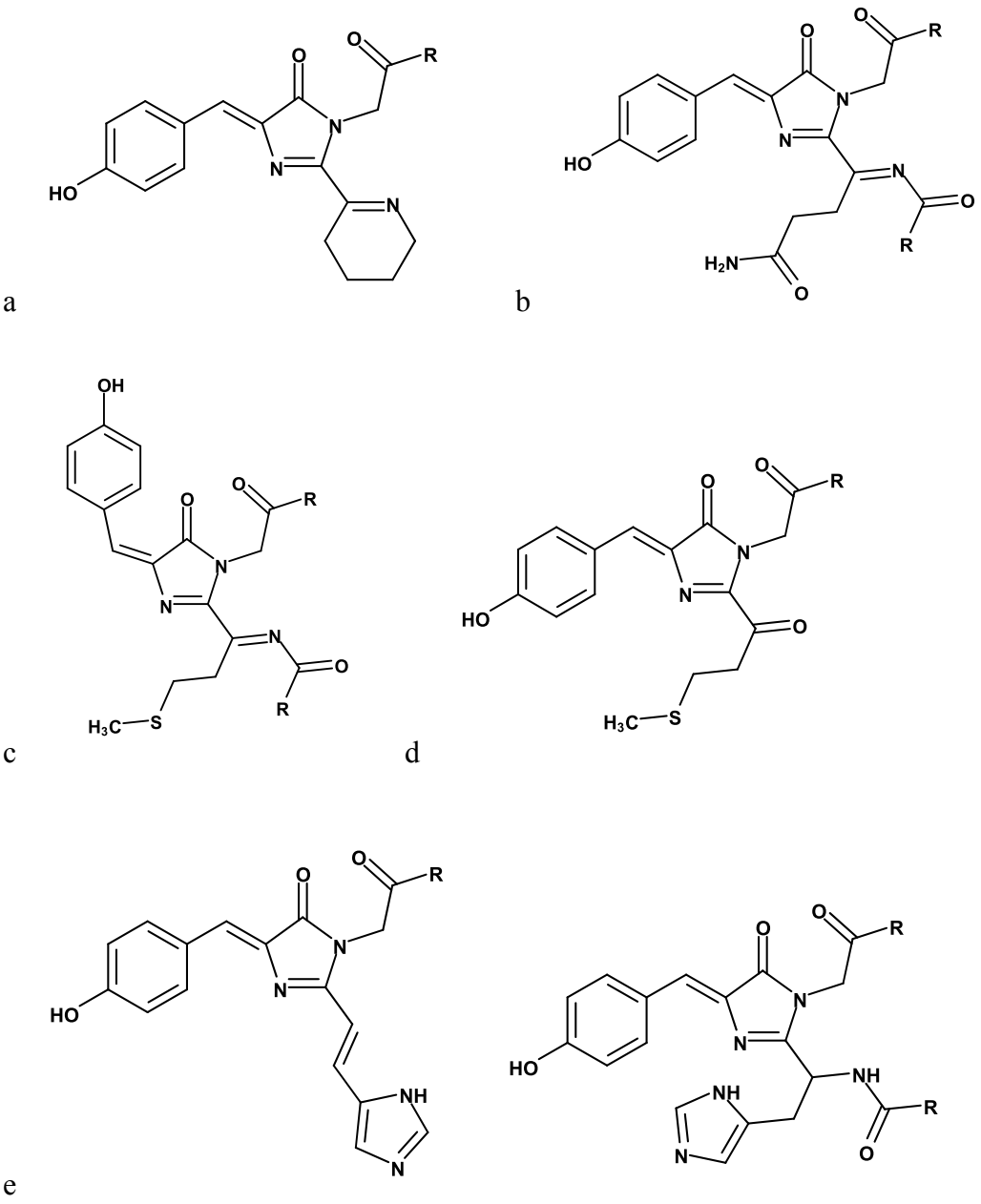


Fig. 3 Structures of chromophores in naturally existed colored FPs. They are zFP538 (a), dsRed (b), eqFP611 (c), asPF595 (d), Kaede, in green fluorescent form (left) and red fluorescent form (right) (e).

In 1999, Matz et al successfully cloned the first red FP (RFP), DsRed, from coral *Discosoma* sp. [27]. Then in 2000, its crystal structure and chromophore structure (Fig. 3b) were solved by Wall et al [28] and Gross et al [29] respectively. Dehydrogenation between α-carbon and nitrogen on Gln66 with catalysis by oxygen forms a N-acylimine structure connecting to the five-member ring cyclized by Tyr67 and Gly68, which introduces two more double bonds to the conjugated system and leads to red-shift of λ_{max} to 583nm [29]. Discovery of this protein later became a ground stone for developing a series of RFP variants commercially available. In 2002, eqFP611, emitting fluorescence in the far-red area on the scale, was isolated from the sea anemone *Entacmaea quadricolor* [30]. A 2.0Å crystal structure shows the far-red light may come from contribution of a *trans*-coplanar chromophore constructed by Met63-Tyr64-Gly65 (Fig. 3c), a parallel ring-stacking interaction with His197, and interaction with other surrounding residues [31]. In 2004, Kusabira Orange (KO), also considered to be the first orange-color-emitting FP, was discovered in the stony coral *Fungia concinna* [32]. Although its crystal structure has never been reported, the monomer variant mKO has been studied (Fig. 4a), and it is believed that Glu212 plays a key role in the formation of a five-membered thioamide ring by Cys65 on the chromophore [33]. In 2007, another RFP from *E. quadricolor*, eqFP578, was cloned and found to share 76% homology in amino acid sequence with eqFP611, including the chromophore, but no crystal structure was solved for it [34]. All these wild-type RFPs are parents of a number of RFP variants that have been applied in different fields.

Despite their construction by β-barrel and chromophore structures, it is hard to consider the GFP-like proteins in purple-blue color as FPs, because they do not emit fluorescence even though absorbance is observed [16]. The chromophore of asPF595, a purple nonfluorescent protein isolated from *Anemonia sulcota*, is formed by the same three amino acid residues to eqFP611, but in a different structure with the backbone cut off, as shown in Fig. 3d [35]. The chemical reason for their non-fluorescence and the biological function are still unknown.

Besides the above, it is worthy to mention that several GFPs display the property of photoconvertibility. Kaede, a GFP from *Trachyphyllia geoffroyi*, has a three-ring chromophore formed by His62-Tyr63-Gly64 (Fig. 3e), which can switch green fluorescence to red after exposing to UV light. It is achieved by a β-elimination reaction on N^{α}–C^{α} of His62 after being excited by UV and releasing the water molecule binding to the His62 imidazole ring, which

prolongs the π-conjugated system and leads to red-shifting [36, 37]. EosFP from *Lobophyllia hemprichii* [38], Dendra from *Dendronephthya* sp. [39] and KikG from *Favia favia* [40] also show similar ability under UV or blue illumination. The case of Dronpa, a GFP isolated from *Pectiniidae* sp., is a little bit different because it changes between fluorescence and non-fluorescence through irradiation with UV and blue lights [41]. This property has not been discovered in FPs with other colors until now.

So far as has been reported, all chromophores in wild-type FPs are generated by autocatalytic modification between three continuous amino acids in a Xxx-Tyr-Gly template, as shown in Table 1. The cyclization between Tyr and Gly makes them share a same imidazolinone core and a mutual π-conjugated system with excited-state energy levels located in the visible light spectrum [42], while the identity of the first amino acid plays a dominant role in the color of fluorescence produced. The chemical properties of the side chain and the further modification after the imidazolinone is formed may increase the length of the conjugation chain, pushing the excitation and emission wavelengths to the red area of the visible spectrum. One more double bond on the GFP chromophore would convert it into YFP. Meanwhile, interaction between the chromophore and amino acid residues on the β-barrel, water molecules and/or solvent molecules, may greatly influence the fluorescence as well. With an identical chromophore structure under different chemical environments, amFP486, dsFP483 and cFP484 emit cyan light when avGFP emits green, KO looks orange while Dronpa switches between green and red.

Table 1 Selected FPs and the amino acid residues forming chromophores.

Selected FPs	Residues forming chromophores
avGFP, amFP486, dsFP483, cFP484	Ser-Tyr-Gly
zFP538	Lys-Tyr-Gly
DsRed	Gln-Tyr-Gly
eqFP578, eqFP611, asFP595	Met-Tyr-Gly
KO, Dronpa	Cys-Tyr-Gly
Kaede, EosFP	His-Tyr-Gly

2. Development of FP variants

It was fortunate to discover avGFP first, because it has multiple strengths favoring the applications for scientists. It is non-toxic to any organism, easily fused to proteins, expressed and tracked *in vivo*. Also, it naturally forms into a weak dimer with a small mass (~50kD), allowing the avGFP-labeled protein to keep the original structure and function. Its ease-of-use helped the FP technology to develop with great interest. However, unlike avGFP, most wild-type FPs exist as obligate tetramers, whose large size prohibits them to be fusion tags for proteins. Some of them show poor brightness and/or stability *in vivo*, and are found lethal to living cells. Moreover, the issues with slow folding and incomplete formation of chromophores commonly exist in wild-type FPs. Baird et al reported that DsRed took about 27 hours to reach half its maximum red fluorescence at room temperature, and more than 48 hours to be maximum; a certain amount of green fluorescence was detected during the formation of the chromophore, implying an incomplete maturation [43]. Considering the huge potential in FP application as research tools, scientists are making efforts to overcome the limitations through developing FP variants with multiple bioengineering methods including random mutagenesis by chemical mutagens, error-prone PCR, deliberate site-directed mutations, randomization of a predetermined stretch of amino acids, DNA shuffling, heteroduplex recombination, directed evolution [12].

Another motivation to develop FP variants is that fluorescence emitted by all reported wild-type FPs has maxima in limited wavelength ranges, leaving obvious gaps in the spectrum. Though searching new FPs is still under process, it is time-consuming and sometimes it needs some luck. To not only seek better research tools in a faster manner but also satisfy curiosity from scientists, filling the empty space on the spectrum via manipulating protein structures can be a more accessible direction.

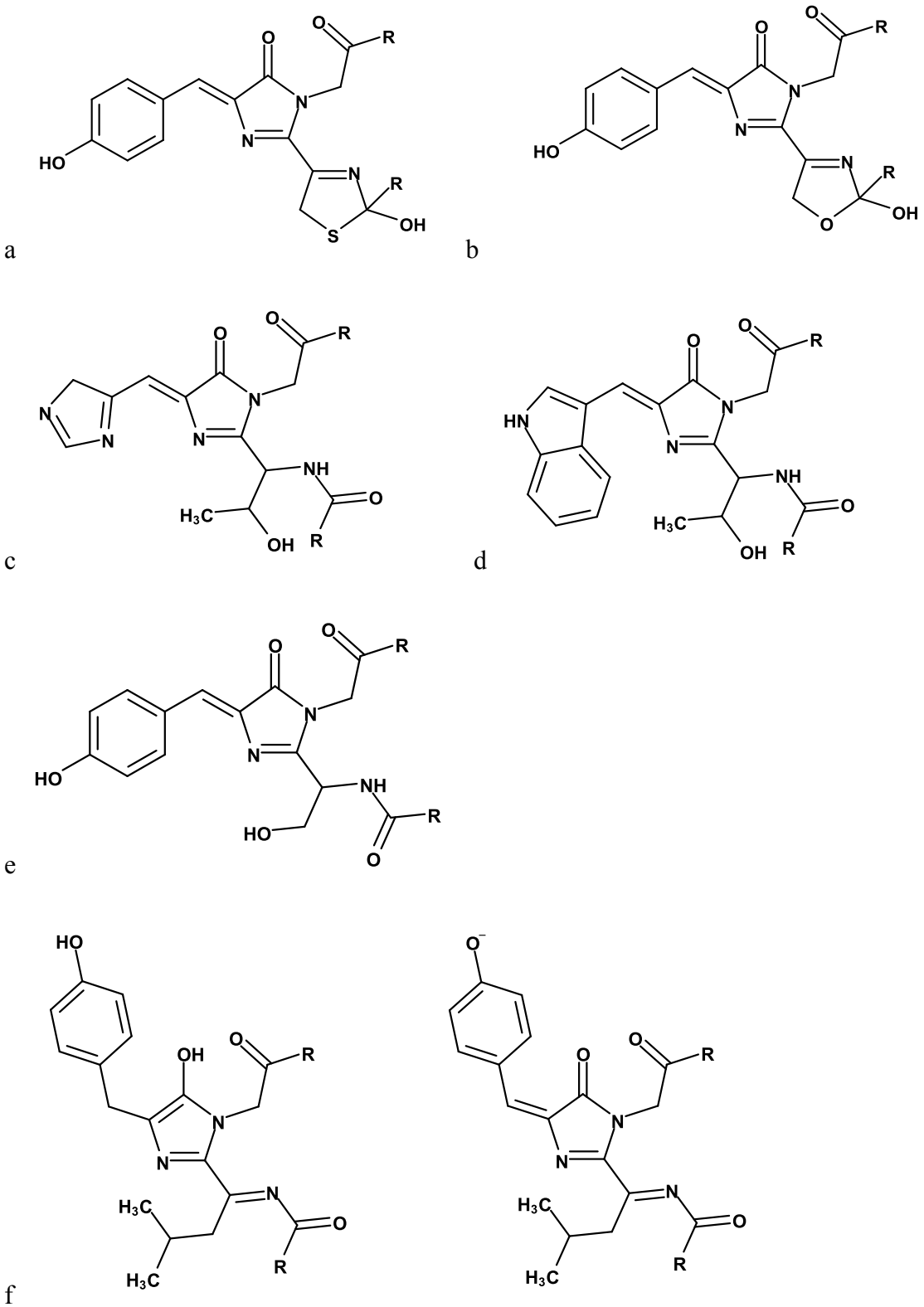


Fig. 4 Structures of chromophores in FP variants. They are mKO (a), mOrange (b), EBFP (c), ECFP (d), EGFP (e), TagBFP (f) in the blue form (left) and the red form (right).

Currently the mutation research mainly focuses on three strategies: introducing amino acids to the positions connecting molecules to interrupt formation of tetramer, manipulating the chromophore structure and/or the residues on the β-barrel to improve quality of FPs for favoring the application. They may be applied solely or together. Now a lot of FP variants have been generated.

As the first FP discovered, mutants of avGFP were investigated much earlier than other wild-type FPs. Plenty of variants have been generated, and some of them are serving as important tools in labs. In 1995, Heim et al replaced Ser65 with Thr in the chromophore, and the variant showed strikingly less chromophore-forming time without obvious shift of excitation wavelength [44]. To keep this advantage and increase the brightness, mutation F64L was further executed by Cormack et al, and EGFP emitting 35-fold brighter fluorescence than avGFP was created (Fig. 4e) [45]. With further substitution of Thr203 to Tyr or His, the protein emitted yellow fluorescence, probably because the polar aromatic residues could lower the excited-state energy of the adjacent chromophore [46]. When mutation happened to Tyr66 instead of Ser65, the change was also dramatic. The exchange of Tyr66 to Trp introduced an indole in the chromophore, blue-shifting λ_{max} from 508nm to 480nm with weak cyan fluorescence produced[47]. The further mutation of F64L/S65T/N146I/M153T/V163A generated a brighter version, ECFP (Fig. 4d) [48]. Yet the replacement of Tyr66 with His could move the emission wavelength to further blue area with λ_{max} at 448nm, emitting weak blue light [47]. With replacement of F64L and Y145F, EBFP was created with improved brightness [48, 49]. Because of their dim color and being easily photobleached, blue fluorescent proteins (BFPs) are less practical than other FP variants. Although the research is still ongoing, scientists realize that it is hard for GFPs to evolve into variants beyond yellow-green. Therefore, to produce red FP variants, the original species needed to be screened for RFPs.

Luckily, a number of wild-type RFPs has been discovered and isolated. Between them, DsRed is the representative mother for generating a family of mutants, as shown in Fig. 5. Obtained from mutation experiments for improving efficiency of chromophore formation [50], a DsRed variant DsRed.T1 was chosen as the starting point for developing dimeric or monomeric mutants. Campbell et al successfully interrupted formation of the tetrameric structure through inserting Arg with positive charge at the positions central to the interface between subunits to

create electrostatic repulsion. With a total of 33 amino acids being substituted, the first monomeric RFP, mRFP1, was generated [51]. It was then quickly adopted by scientists as a fusion tag in multiple applications. However, its incomplete formation of chromophore, unpredictable brightness and poor photostability for long-time imaging reduced its reliability. To overcome these weaknesses and to produce FPs emitting different fluorescence, a lot of efforts was done to improve mRFP1, including exchanging Gln66 and/or Tyr67 on the chromophore with Met, Cys, Thr and/or Trp, and substitution of residues surrounding the chromophore. Many high-performing YFPs and RFPs widely adopted today were developed, as shown in Fig. 6, and they are now known as "mFruits". Specially, mApple is the parent of a RFP variant, K78BoF, directly involved in my project, so more details about mApple is provided later.

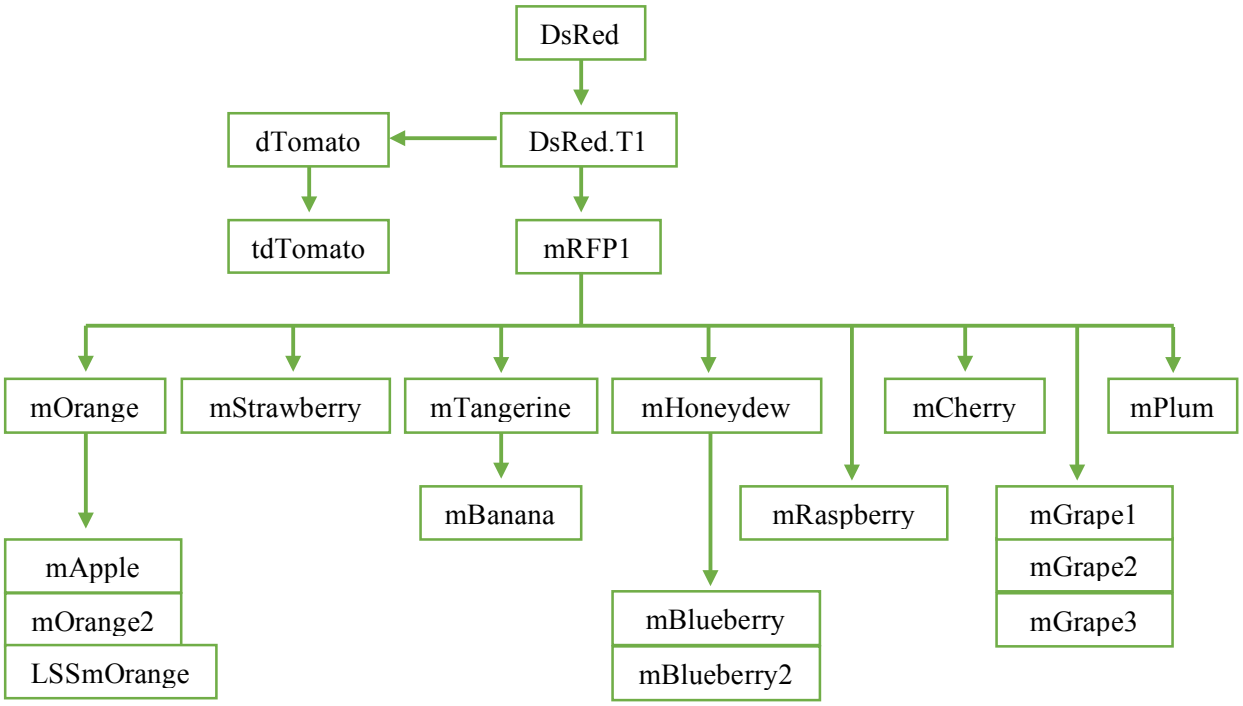


Fig. 5 Flowchart for the development of FP variants derived from DsRed [42, 52].

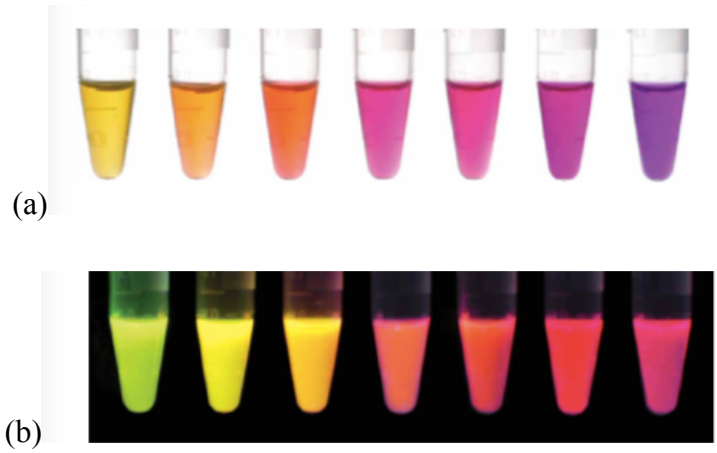


Fig. 6 Some purified DsRed-derived FP variants shown in visible light (a) and fluorescence (b) [52]. From left to right, the order of the proteins is as follows: mHoneydew, mBanana, mOrange, tdTomato, mTangerine, mStrawberry, mCherry.

The other classical case for developing a system of FP variants comes from *E. quadricolor*, in which two RFPs eqFP578 and eqFP611 were isolated and mutated into diverse high-efficient variants, as shown in Fig. 7. The first monomeric variant from *E. quadricolor*, TagRFP, was derived from eqFP578 through many rounds of random or semi-random mutagenesis, which was 2.8-folder-brighter than mCherry and considered to be the brightest at that time [34]. Then it was rationally mutated into TagRFP-T through a substitution of S158T, and is one of the most photostable monomeric FPs currently [53]. Especially interesting is the fact that TagRFP could even be converted into a bright monomeric BFP, mTagBFP, with λ_{max} at 456nm through a replacement of Met63 to Leu and mutations on several residues surrounding. It creates a completely different chromophore structure to other BFPs, as shown in Fig. 4f [54, 55]. Meanwhile, another direction for mutating eqFP578 was conducted to screen out variants with long emission wavelength. Katushka was selected with λ_{max} at 635nm and further optimization on it resulted in several far-red FP variants created, including mNeptune, a far-red FP three-fold brighter than mPlum at 650nm [56, 57]. Although eqFP611 was discovered first, its variants were successfully generated much later than those of eqFP578. The most representative one is a RFP, mRuby, which is closely related to my project and is discussed in detail later.

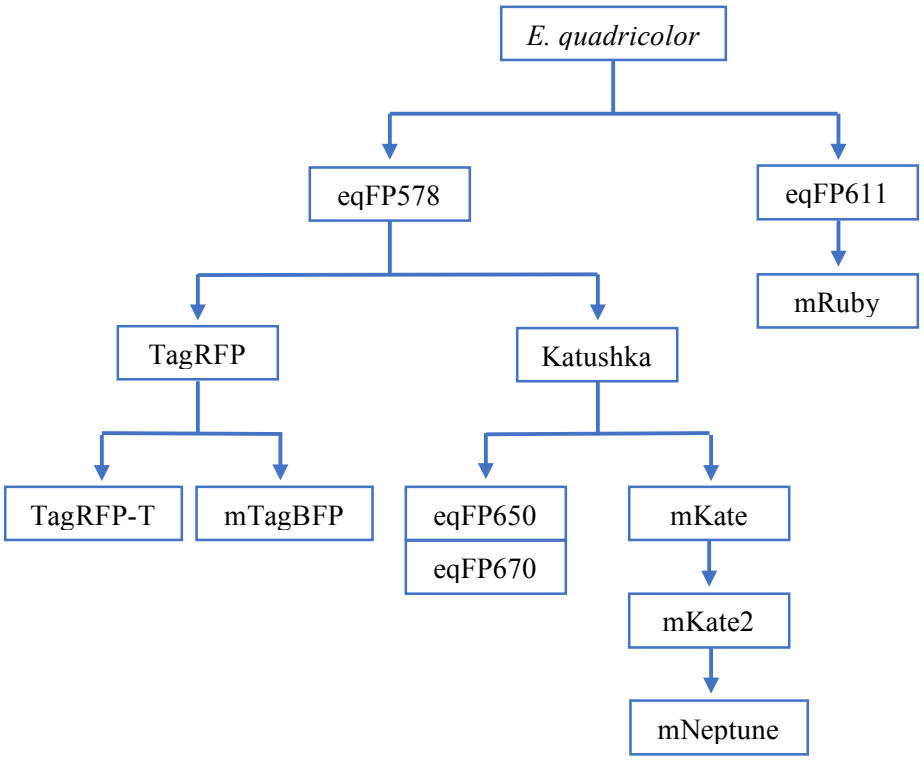


Fig. 7 Flowchart for the development of FP variants derived from eqFP578 and eqFP611, two RFPs isolated from *E. quadricolor* [42].

3. Application of FPs or FP variants

Ever since Chalfie et al proved that avGFP could be expressed in eukaryotic cells without cytotoxicity or losing the green fluorescence [10], FPs were widely applied to monitor the activation of promoters or expression of proteins through inserting FP genes after promoter regions or fused to the target genes [13]. Later, multiple FP-based biosensors were developed for experimentally reading out molecular interactions between a FP-labeled protein and an analyte of interest which could be a small molecule, a protein or an enzymatic activity [58]. Now FP-based biosensing is one of the most important applications of FPs.

Depending on the design and purpose of an experiment, one FP or two may be employed. Single FP-based biosensors are usually composed of a premature FP and an exogenous or

endogenous protein. When a small molecule or an ion binds to the FP or fused protein, the conformation of the chromophore will switch into the active state and fluorescence can be detected. Biosensors in this kind have successfully measured concentration of Ca^{2+}, Zn^{2+}, ATP, cGMP in vivo [58]. Two FPs-based biosensors, commonly known as Förster resonance energy transfer (FRET)-based biosensors, are particularly valuable in analyzing enzyme activity and protein-protein interactions *in vivo*. In FRET between two FPs, a nonradioactive energy transfers from an excited donor chromophore to an acceptor chromophore in a distance less than 10nm, triggering the acceptor emitting fluorescence. When a protease splits the substrate labeled by two FPs with different colors on each terminal, fluorescence of the acceptor will be diminished because of the increasing distance. Alternatively, interaction between two proteins tagged with FPs can be traced through detecting the light emission from the acceptor, which was used to study the oligomerization state of members in the G-protein-coupled-receptor superfamily.

Actually, research in FP applications was more difficult than originally expected. Despite that avGFP shows a lot of strengths in application mentioned before, the weaknesses for it and most of other FPs with short-wavelength spectra are obvious. The strong photobleaching and low transmission rate in cells strikingly reduced the practicability as biosensors. These become motivations for scientists to develop RFP variants in recent years, for long-wavelength red fluorescence scatters less *in vivo* and penetrate tissues better. Also, living cells are less sensitive to longer wavelength excitation and the autofluorescence from cells can be reduced under this condition. Moreover, the development of RFPs can help to achieve the two-color living-cell imaging, which was prevented by the overlapping emission spectra of early variants. Extension of the emission range into the red and even the far-red area helps to better distinguish the emission peaks between different colors [59].

Before RFP was available in the lab, it was CFP-YFP pairs were commonly used in FRET experiments, but they were also problematic because of phototoxicity, photobleaching and photoconversion [60]. Now some GFP-RFP pairs are able to conquer the difficulties [61]. For example, in 2016, mClover3-mRuby3 pair was used with satisfactory brightness and photostability, and successfully employed to improve response in the calcium/calmodulin-dependent kinase II alpha (CaMKIIα) reporter [62].

4. Two RFPs in my research

4.1 Protein K78BoF

Dr. Huiwang Ai's lab at University of California, Riverside mainly focuses on developing novel molecular probes by protein engineering, fluorescence and bioluminescence imaging, and synthetic chemistry techniques, to molecularly monitor signal pathways involving redox-active molecules, neurotransmitters, and protein post-translational modifications in the real time. The research efforts are applied to clarify mechanisms of toxicity of chemicals, cancer development and progression, cognition and behavior, and neurological disorders. Recently, they successfully developed a number of FP-based biosensors for imaging mitochondrial membrane, detecting peroxynitrite or Zn^{2+} and monitoring thioredoxin redox activity, and so on. Specifically, a GFP-based biosensor, pnGFP, was generated by introducing an unnatural amino acid, *p*-boronophenylalanine (BoF) (Fig. 8), to selectively detect production of peroxynitrite *in vivo*, a redox signaling molecular related to some serious diseases including various cancers and Alzheimer's disease [63].

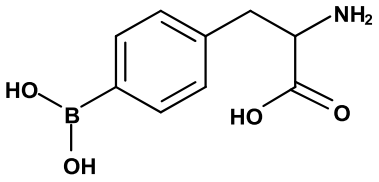


Fig. 8 Chemical structure of *p*-boronophenylalanine (BoF).

Protein K78BoF, provided by the Ai lab in this project, is an RFP-based sensor used to sense peroxynitrite and hydrogen peroxide. It was improved from circularly permuted mApple by substituting Lys30 with BoF. As an ideal parent protein, mApple was derived from mOrange, a variant of wild-type RFP DsRed (Fig. 5), by mutation on the chromophore (T66M) and multiple residues on the β-barrel, resulting in red-shifting to 592nm for λ_{max} and showing brighter fluorescence than mCherry [53]. The experimental results from the Ai lab show that the further mutation of mApple by BoF offers K78BoF a biosensing function, but the protein

becomes sensitive to the change of pH at the same time. When the pH value is lower than approximately 7.5, it loses its red color and visually turns to yellow. Further study by crystal structure determination may assist the Ai lab to clarify the possible reason.

4.2 Protein mRuby0.4-3

Dr. Jun Chu's lab at Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences was founded in 2016. His current research interests include developing new single RFP-based biosensors for living cell imaging and improving FPET-based biosensors for monitoring *in vivo* activity of kinases.

As a novel RFP-based biosensor developed from mRuby3, mRuby0.4-3 was provided a gene inserted in a plasmid by the Chu lab.

mRuby was derived from eqFP611 through mutation of F102I, insertion of Arg122 and Ala194, and the replacement of C-terminal sequence and amino acids that prevented chromophore folding correctly [64]. Its chromophore can be formed in several hours with unusually bright red fluorescence emitted at 605nm, which fills the spectra gap between green/yellow and far-red FPs with very high extinction coefficient but low quantum yield. It also shows strong resistance to denaturation in the broad range of pH. Because of the advantages it displays, it was further engineered to improve brightness, maturation time, and photostability. The second and third generation mRuby2 and mRuby3 were developed and successfully applied in GFP-RFP FRET experiments [60, 62]. Furthermore, mRuby3 is currently one of the brightest monomeric RFP tags. As the daughter of mRuby3, mRuby0.4-3 maintains similar strengths, but also shows tolerance to a narrow pH range from 6.0 to 7.0 for fluorescence emission. The red fluorescence is diminished significantly when pH is lower than 6.0 or higher than 7.0. To explain this phenomenon, understanding how the conformation of the protein change at different pH values is essential.

RESEARCH GOALS

1. Crystal structure determination of K78BoF

This project is to identify the effect of *p*-boronophenylalanine (BoF) substituting for Lys30 on fluorescence of the protein through analyzing the crystal structure of K78BoF in a weakly acidic or neutral environment, which will provide data support to the further development of K78BoF on research.

2. Crystal structure determination of mRuby0.4-3

Two goals are expected in this project. The first one is to decipher the 3D structure of this new red FP variant. The second goal is to describe the effect of pH change on the conformation, given that its fluorescence is pH-sensitive.

MATERIALS AND METHODS

1. General scheme

To obtain a crystal structure of a protein, four steps of operation are usually required, as shown in Fig. 9. Firstly, the protein is overexpressed in *E. coli* cells transformed with the protein-gene-inserted plasmid. Then the cells are harvested and lysed, followed with nickel-affinity column and size exclusion chromatography purifications sequentially. When the protein is pure enough for crystallization, preliminary screening of crystallization conditions is conducted to narrow down the potential conditions. Depending on the result of the screening, further optimization of the selected methods may be necessary to gain crystals with good quality. The crystals collected are shot with X-rays, and diffraction patterns are recorded for data analysis and solving the 3D structure.

Protein K78BoF was prepared by Ai's lab and sent to Ng's lab in a purified situation, so it was ready for crystallization experiments. Protein mRuby0.4-3 was provided as a gene inserted into a plasmid by Chu's lab, thus my research started from the first step. Expression of protein and purity of proteins in each purification step were checked by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

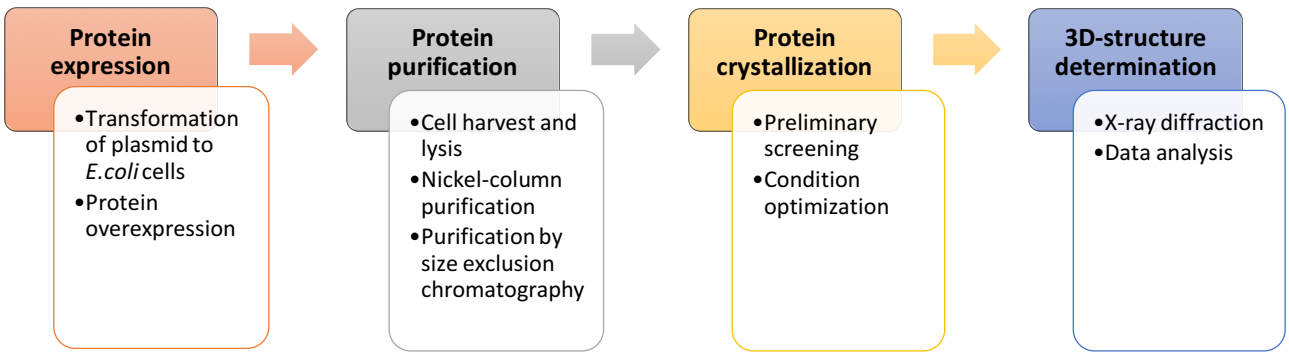


Fig. 9 General scheme for 3D-structure determination of proteins.

2. Techniques applied in protein expression and purification

2.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a technique used to analyze proteins or peptides based on their weight difference in an electrical field. As an anionic detergent, SDS disrupts the secondary, tertiary and quaternary structures of proteins and turns them into linear polypeptide chains. Meanwhile, it binds to the chains in a ratio of approximately one SDS molecule to two amino acids, making them negatively charged proportionally to protein mass. The technique is able to separate proteins with molecular weight from 5 to 2,000 kDa. When subjected to an electric field in PAGE, the peptide chains move toward the anode at different rates inversely related to their weight. The denatured proteins can be then separated and analyzed through dyeing the gel with Coomassie blue.

Mini-PROTEIN® TGX^{TM} Gels purchased from Bio-Rad Laboratories, Inc. with acrylamide percentage in 8-16% were applied to check the efficiencies of protein expression and purities of proteins here. The samples were mixed with loading buffer and loaded into the wells of the gel. The gel was submerged in tris-glycine buffer and run in an electrical field under no more than 200V. Then the gel was rinsed by DI water three times with 15 minutes each time, and stained in Coomassie blue under room temperature for about one hour. After rinsing it by DI water once, the gel was ready for analysis.

2.2 Nickel-column purification

Because the six histidine residues tagged onto the amino acid sequence of mRuby0.4-3 show high affinity to immobilized nickel ions (Ni2+), columns filled with Ni2+-immobilized resin can selectively retain the tagged proteins while other proteins or contaminants are washed away with buffer. Then the proteins can be gently eluted through competition by high concentration of imidazole 64. HisTrapTM FF crude column prepacked with 1mL of precharged Ni Sepharose™ 6 Fast Flow, produced by GE Healthcare, was utilized in the purification step of mRuby0.4-3. Solutions used are listed in Table 2.

After washing the column with 6mL of DI water, I equilibrated it with 6mL of washing buffer. Then I loaded the sample into the column and collected the solution eluted. The column was washed by 15mL of washing buffer for removing the contaminants not binding to it, while the eluted liquid was collected as well. After that, I injected 6mL of elution buffer into the column and collected the eluate into different tubes for every 1mL. The column was regenerated through being rinsed by a series of solutions with a syringe, in an order as follows: 10mL of stripping buffer, 6mL of 8M urea, 5mL of binding buffer, 5mL of DI water, 30mL of cleaning solution, 10mL of binding buffer, 10mL of DI water, 1mL of charging solution, 5mL of DI water, 5mL of binding buffer.

Table 2 Components of solutions applied in nickel-column purification.

Solution	Component (with concentration and pH)
Striping buffer	20mM Sodium Phosphate, pH 7.4 *
	0.5M Sodium Chloride
	50mM EDTA
Binding buffer	20mM Sodium Phosphate, pH 7.4 *
	0.5M Sodium Chloride
Cleaning solution	1M NaOH
	1.5M NaCl
Charging solution	0.1M $NiSO_4$
Washing buffer	20mM Imidazole
	20mM Tris, pH 8.0
Elution buffer	200mM Imidazole
	20mM Tris, pH 8.0

* To prepare 20mM Sodium Phosphate, pH 7.4, Henderson-Hasselbach equation, pH = pKa + log ($[Na_2HPO_4]/[NaH_2PO_4]$), with pKa of 6.9 was employed to calculate the mass of Na_2HPO_4 and NaH_2PO_4.

2.3 Size exclusion chromatography

Size exclusion chromatography (SEC) is a method for separating macromolecules such as proteins in a solution by molecule size. Usually it is performed with a column filled with porous polymer beads. When solutions travel down the column, smaller molecules are trapped in the pores while the larger molecules pass between the beads and elute faster. Here mRuby0.4-3 solution was purified through prepacked SuperdexTM 75 10/300 GL column using an ÄKTA protein purification system from GE Healthcare Life Sciences. To prepare the column before purification, it was washed with sterilized DI water for 2h and then equilibrated by elution buffer (20Mm Tris, 100Mm NaCl, pH 7.5) for about 1h at a flow speed of 0.6mL/min. After sample solution was loaded into the column, elution buffer was applied as mobile phase to separate proteins at a flow speed of 5.0mL/min. The eluate was collected into different test tubes for every 2 minutes controlled by a fraction collector. When purification was completed, the column was washed by sterilized DI water for 2h.

3. Protein crystallization

3.1 Preliminary crystallization method screening

MCSG (Midwest Center for Structural Genomics) crystallization suites purchased from Anatrace were utilized for preliminary crystallization method screening. There are four screen suites available in Ng's lab with 96 different crystallization conditions in each one, which are MCSG-1, MCSG-2, MCSG-3 and MCSG-4. Each condition contains salt, buffer and/or precipitant. Conditions 1-288 in MCSG-1, MCSG-2, MCSG-3 are arranged in a decreasing order of crystal productivity according to results of trials of more than 40,000 diverse proteins, while conditions in MCSG-4 are rationally selected for proteins requiring additional screening.

The Crystal Gryphon (Fig. 10) produced by Art Robbins Instruments, LLC was employed to distribute protein solutions and crystallization reagents in MCSG suites on the wells of Intelli-Plate 96-3 LVR crystallization plates (Fig. 11, purchased from Hampton Research) in three different volume ratios between two solutions as shown on Table 3.

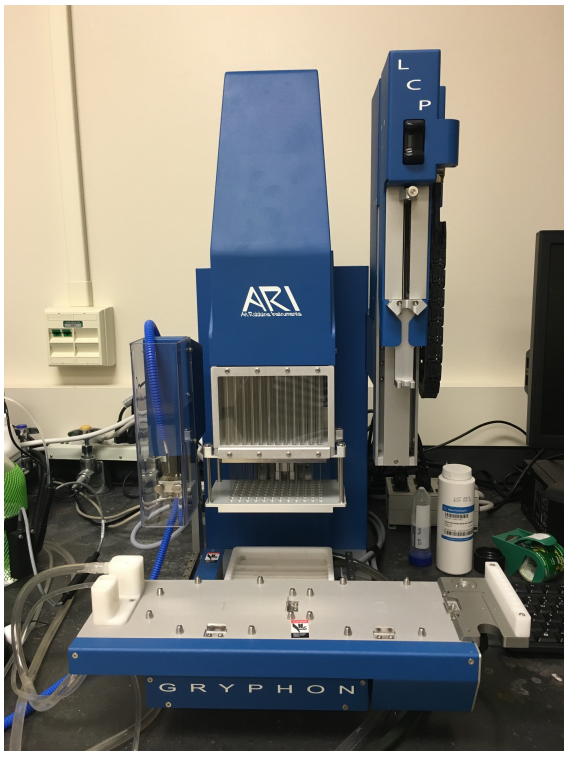


Fig. 10 Picture of Crystal Gryphon.

A. 

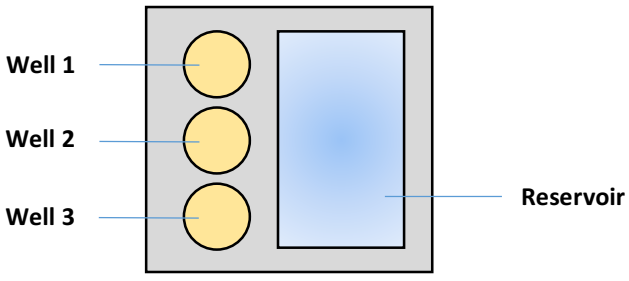
B. 

Fig. 11 Intelli-Plate 96-3 LVR crystallization plate (A) and well arrangement for each condition (B).

Table 3. Volume of protein solution and crystallization reagent placed on each well of Intelli-Plate 96-3 LVR crystallization plate.

Well	$V_{protein}$ (μL)	$V_{reagent}$ (μL)
1	0.375	0.125
2	0.250	0.250
3	0.125	0.375

3.2 Microseeding experiments

In a crystallization experiment, addition of a crystallization reagent to the protein solution can increase the relative supersaturation of the protein which favors nuclei formation and then crystal growth. According to the difficulty of crystal nucleation and growth, supersaturation can be divided into three stages, metastable zone, labile zone and precipitation zone sequentially (Fig. 12). In the metastable zone, nuclei cannot form but crystal growth from seeds can occur, while spontaneous nucleation and crystallization can happen in the labile zone, and protein will precipitate in the precipitation zone. This means that spontaneous nucleation requires higher degrees of supersaturation of the protein solution than crystal growth, to make sure the nuclei can reach a critical size to grow into a crystal rather than redissolve. However, as a technique widely employed to bypass spontaneous nucleation, seeding can assist to obtain crystals with bigger size and better quality. There are mainly two ways to conduct a microseeding experiment, streak seeding and solution microseeding [66]. Because of its easier operation and controllability in an experiment, I chose to perform solution microseeding.

Solution microseeding experiment requires preparation of a seed stock solution via using the Teflon Seed Bead kit (Hampton Research). Protein crystals were picked under microscope and transferred into an iced 1.5mL-centrifuge tube with a bead (Fig. 13). Then the related crystallization reagent was added to the tube to make the total volume to be approximately 50μL. The tube was vortexed for three minutes with a pause on ice for every 30 seconds. The seed stock was then ready and stored at -80C°. Different dilutions of the stock were also prepared based on the experiments and stored at -80C°.

In a microseeding experiment, protein solution, crystallization reagent and seed solution were mixed in fixed volume ratios which depended on the crystallization condition and the experiment design.

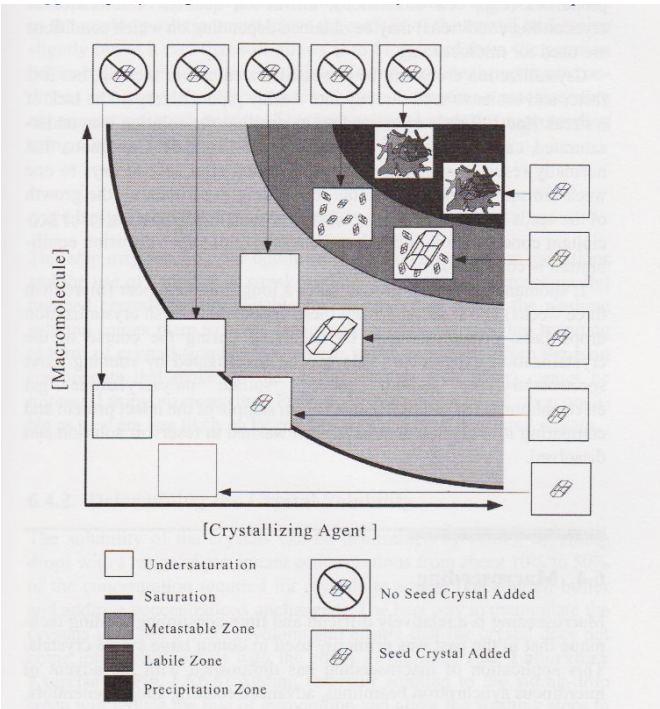


Fig. 12 Seeding and phase diagram [66].

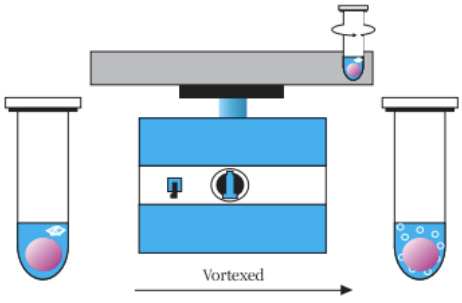


Fig. 13 Seed bead (pink ball) and operation for preparing seed stock solution.

3.3 Additive screening experiment

The Additive Screen HT ™ kit produced by Hampton Research contains 96 wells of different solutions, including multivalent cations, salts, amino acid, dissociating agents, linkers, polyamines, chaotropes, co-factors, reducing agents, polymers, chelating agent, carbohydrates, polyols, non-detergents, amphiphiles, detergents, osmolyte, organic (non-volatile) or organic (volatile) reagents. By adding to conditions in the MCSG suites, it will adjust the environment of protein crystallization including altering pH of crystallization reagents, vapor rate of water, etc. The results may facilitate or diminish the crystallization.

Additive screening experiment was conducted using the Crystal Gryphon to mix the solutions with protein solutions and crystallization reagents in fixed volume ratios.

4. Protein crystal structure determination

Crystals were picked from wells using cryoloops under microscope, rinsed by cryo-protectant and stored in liquid nitrogen, then sent to the Advanced Light Source at Lawrence Berkeley Lab at University of California, Berkeley for collection of single-crystal monochromatic diffraction patterns through using the SIBYLS beamline there. X-ray Detector Software (XDS) [66] was employed to convert the patterns into structure factors and then Phaser [67] was used to perform molecular replacement. The structure model was refined and rebuilt manually by Coot 0.8.9-pre EL [68] and computationally by Refmac5 in CCP4 [69]. Crystal structures of K78BoF and mRuby0.4-3 were performed with The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

RESULTS

1. Crystallization of K78BoF

1.1 Purity of K78BoF

Because the purity of a protein plays an important role in crystal growth and quality, a SDS-PAGE gel was run to check the purity of K78BoF when it arrived. The gel result in Fig. 14A showed that K78BoF was the majority of the protein solution, although small amounts of other proteins were found. The solution was pure enough for protein crystallization.

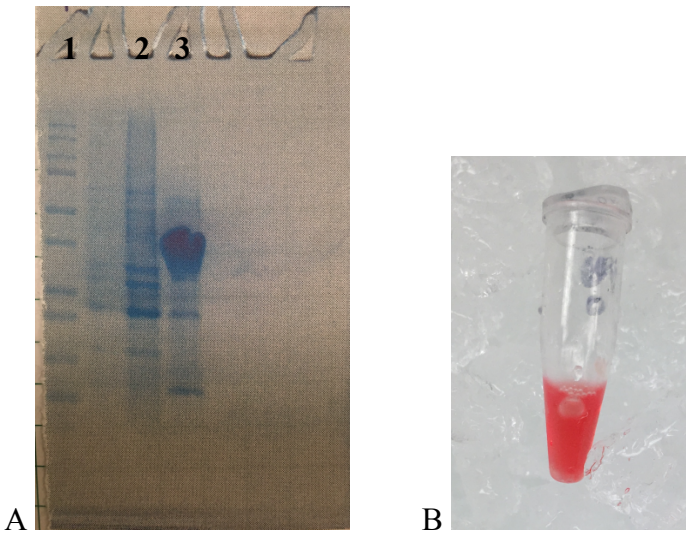


Fig. 14 SDS-PAGE gel result of K78BoF (A) and protein solution in 20mg/mL (B). On A, lane 1, 2 and 3 were for prestained protein ladder, protein S62BoF, and K78BoF respectively. S62BoF was another red FP protein sent with K78BoF planning for crystal structure determination. However, the purity of S62BoF was not good enough for further study, so it was not mentioned in the project.

1.2 Crystallization of K78BoF

1.2.1 Crystallizing method screening

To find suitable conditions for crystallizing K78BoF, preliminary screening was conducted with MCSG-1 and -2 suites in 20mg/mL of K78BoF at 13℃ for two weeks. Three potential conditions, MCSG-1: G10, MCSG-1: G11, and MCSG-2: D12, were identified as seen in Fig. 15 and Table 4. In MCSG-1: G10 condition, spherulites were formed with identical color of the protein solution. Spherulites are transparent, droplet-like solids showing dark and light parts under the polarizer, whose growth is a good starting point for condition optimization [66]. In both MCSG-1: G11 and MCSG-2: D12 conditions, yellowish crystals were formed. Since K78BoF will change color from red to yellow in an acidic environment, the crystals formed should be K78BoF instead of salt. Although both are yellow, needle-type crystals, the ones in the MCSG-1: G11 condition were apparently thicker and clearer in shape, favoring X-ray diffraction experiments. Therefore, I chose MCSG-1: G10 and MCSG-1: G11 conditions for the following optimization experiments.

Table 4 Chemical components for potential crystallization conditions of K78BoF.

Condition	Salt	Buffer	Precipitant	$V_{protein}$: $V_{reagent}$	Crystal Type*
MCSG-1: G10	0.1M magnesium formate		15%(w/v) PEG 3350	1:1	Spherulites
MCSG-1: G11		0.1M Na_2HPO_4: Citric acid pH 4.2	40%(v/v) PEG 300	1:3	1D
MCSG-2: D12	0.2M ammonium sulfate		20%(w/v) PEG 3350	1:1	1D

* According to the shape of crystals, they usually can be divided into three different types, 1D crystals for spine/needle-like shape, 2D crystals for thin plate shape, 3D crystals for chunk shape.

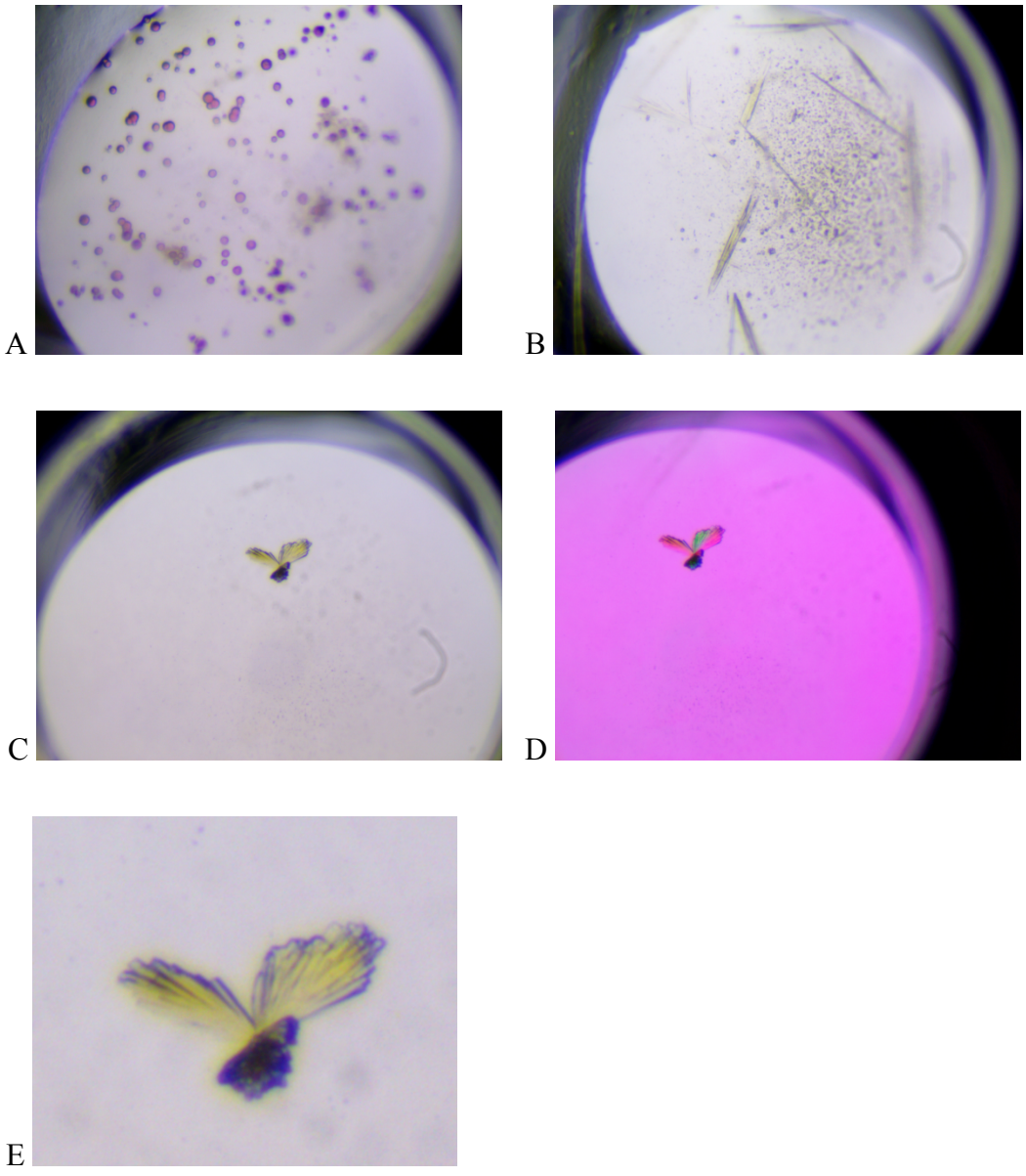


Fig. 15 Potential conditions for crystallizing K78BoF at 20mg/mL. (A) MCSG-1: G10 condition. (B) Crystal grown in MCSG-2. (C) Crystal grown in MCSG-1. (D) Crystal in MCSG-1: G11 condition shows significant birefringence under polarizer. (E) Crystal grown in MCSG-1 in an amplified view.

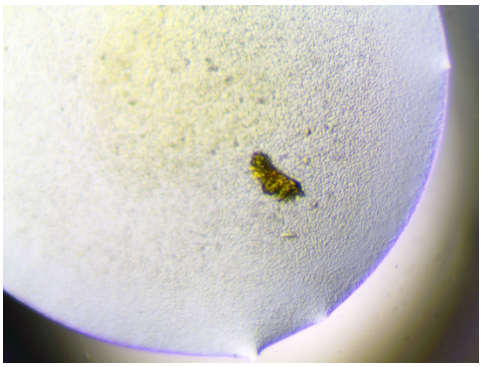
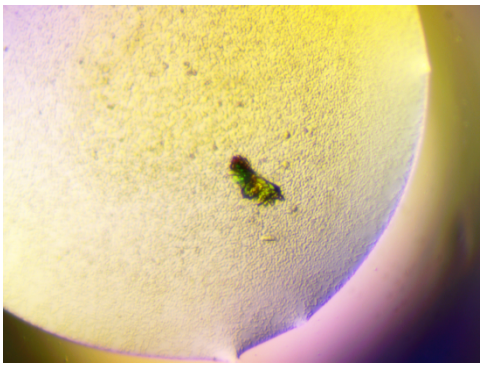
1.2.2 Optimization of MCSG-1: G11 condition

A number of parameters need to be considered to optimize the crystallization method, including protein concentration, pH of crystallization reagent, concentration of salt, buffer and precipitant in the reagent, and volume ratio of protein solution to the reagent. Temperature affects crystal growth as well [66].

In the first run, the effects of K78BoF concentration on growing crystals was explored by increasing the initial protein concentration from 20mg/mL to 55mg/mL. The crystals appeared in a week after mixing protein solution and crystallization reagent, then it took another two weeks to grow larger. Compared to the crystals from 20mg/mL of K78BoF (Fig. 15E), they showed more intensive color, larger size, and thicker shape, indicating that higher concentration favors the generation of the crystals. However, crystals shown on Fig. 15E and Fig. 16C grew in clusters with overlap of other pieces, which might create an issue in X-ray diffraction since it requires a single piece of crystal. Thus, further optimization was needed.

To speed up crystal growing and improve crystal quality, a microseeding method was applied to the following optimization experiments with seed stock prepared from crystals in Fig. 16.

In microseeding experiment I, the effect of protein concentration and seed concentration on crystal growth was studied with experimental design and result shown on Table 5. All nine trials were found growing needle-shape crystals. The amount of crystals was increased with increasing protein concentration, while the size and the brightness of the color were increased with decreasing seed concentration.

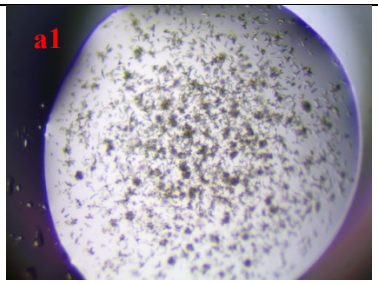
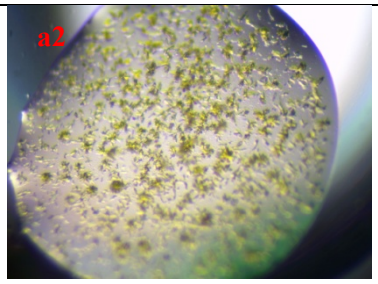
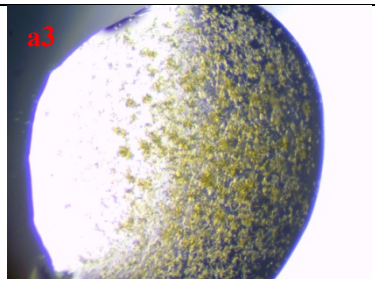
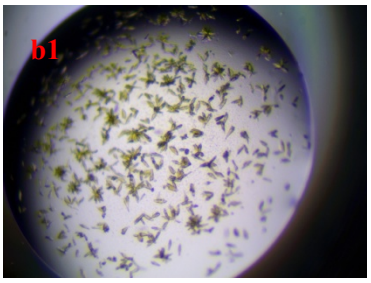
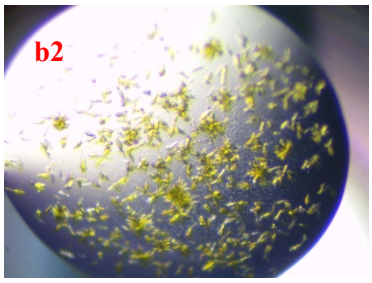
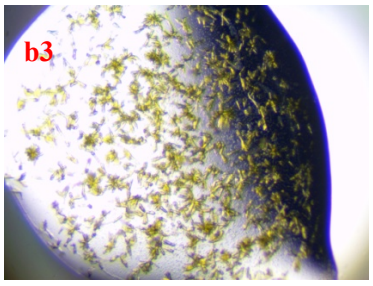
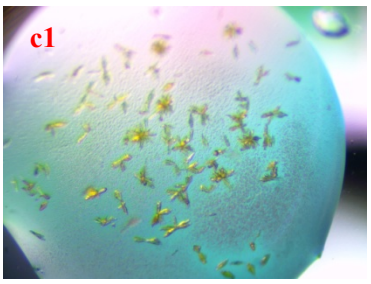
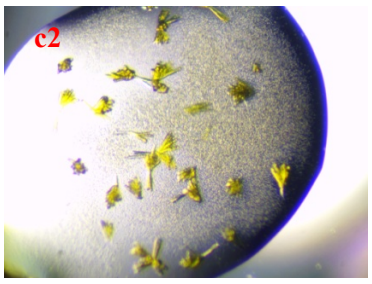
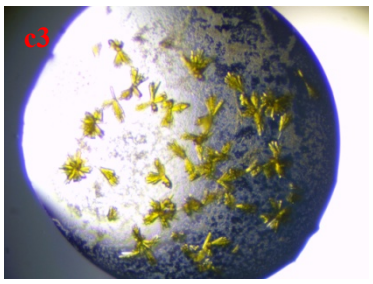
A  B 

C 

Fig. 16 Crystal grown in MCSG-1: G11 condition with initial protein concentration at 55mg/mL. The crystallization happened at 13℃. (A) The crystal was observed without polarizer. (B) The crystal was observed under polarizer. (C) Crystal observed in an amplified view.

Given the brighter color and thicker shape, a number of crystals grown in the adjusted condition b3, c2 and c3 on Table 5 were picked for X-ray diffraction analysis. The diffraction of the crystal from b3 was too weak to be analyzed. The one from c3 was measured to about 2.6Å, which should be good enough to interpret the backbone and most sidechains of K78BoF. Yet the diffraction data could not be collected because several diffraction patterns overlapped with each other, meaning that more than one crystal were shot by X-ray in the same time. Luckily, a single crystal from c2 provided a very clean diffraction pattern with 2.13Å resolution. Its data was collected and ready for 3D-structure determination. However, the MCSG-1: G11 condition crystallized the protein in a relative low pH environment, which was still far from what we expected, meaning more optimization experiments were needed.

Table 5 Microseeding experiment I of K78BoF in MCSG-1: G11 condition with different protein concentrations and seed solutions.

* Seed Stock Dilution	Protein Concentration (mg/mL) **		
	11	22	33
1:1			
1:10			
1:100			

* Seed stock solution was ten-fold diluted by MCSG-1: G11 reagent.

** 55mg/mL of protein solution was diluted to 11mg/mL, 22mg/mL and 33mg/mL by buffer (20mM Tris, pH 8.0).

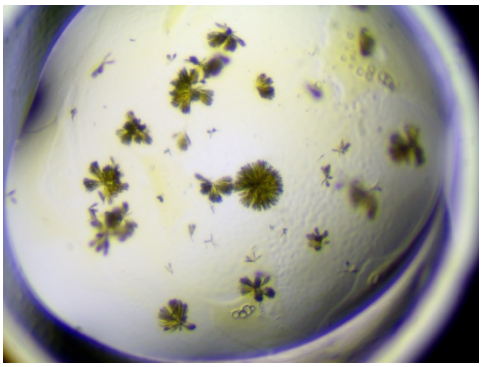
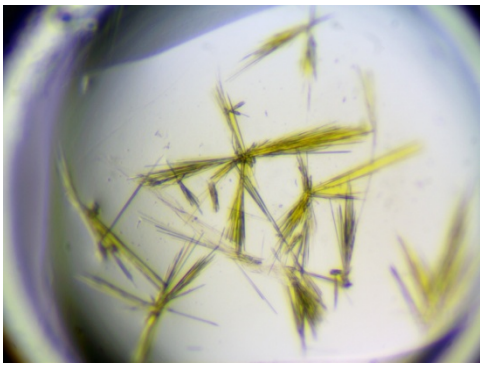
A  B 

Fig. 17 Crystals grown in additive screen combining with microseeding for 20mg/mL of K78BoF in MCSG-1: G11 condition. The additive screen was conducted in 13℃° with a volume ratio of solutions as follows: (protein solution): (crystallization reagent): (additive screen reagent): (hundred-fold diluted seed stock) = 5: 4: 1: 1. (A) Additive screen-A2 condition: 0.1M Cadmium chloride hydrate. (B) Additive screen-B7 condition: 0.5M sodium fluoride.

An additive screen combined with microseeding was conducted to improve the possibility of crystal growth. After incubating the plate for 6 days, only two conditions were found to grow yellowish crystals (Fig. 17). These conditions showed thinner needle shapes than the previous crystals, indicating the adjustments created by reagents in additive screen could not facilitate K78BoF to form crystals with higher quality.

Given the unexpected results from the additive screen experiment, focus on optimization of MCSG-1: G11 condition itself seemed to be a better direction to obtain higher-resolution crystals. Thus, more microseeding experiments based on MCSG-1: G11 condition were developed.

Microseeding experiment II was designed to explore the effects of pH value of the buffer and types of precipitants on crystal growth for 10mg/mL and 20mg/mL of K78BoF at 4℃° (Table 6). Multiple aspects were considered when I designed it. As the goal of this project is to solve K78BoF structure at a pH close to 7.0, increasing the pH value became necessary. Being a commonly used precipitant, PEG is able to induce protein–protein attractive interactions to form precipitates including crystals. Its molecular weight and concentration also influence the

crystallization of proteins [70]. Thus, PEGs with different molecular weight and concentration were introduced to the experiment. Also, the temperature can affect the vapor rate of water molecules, which will further affect the growing speed of crystals. Lowering the temperature can help to form crystals with slower rate but better quality.

After incubating for approximately two weeks, crystals were generated in the conditions labelled on Table 6. Generally, low protein concentration and pH value favored crystal growth, but crystals could be grown in high pH, implying the possibility of achieving that goal. However, the crystals were grown in clusters with needle-shapes (Fig. 18A) or in lumps (Fig. 18B), which created problems in the X-ray diffraction experiments. Although several plate-shape crystals were found (Fig. C), they were too thin for analysis by X-ray diffraction.

Many crystals were picked from different conditions for X-ray shooting, but none of them could provide clear enough diffraction pattern for 3D-structure determination. The patterns were either cracked or smeared, meaning the crystals overlapped with each other or the crystal structure was partially disordered. The growth condition or the cryo-protectant was not suitable to the protein.

Table 6 24-well plate arrangement for microseeding experiment II with 0.1M Na_2HPO_4 and 1000-fold diluted seed stock at 4℃.

pH	35%(v/v) PEG 400	35%(v/v) PEG 500	30%(w/v) PEG 1000	20%(w/v) PEG 1000
4.2	A1*△	A2*△	A3*△	A4*
5.0	B1*△	B2*△	B3*△	B4
6.0	C1*△	C2*	C3*△	C4
8.0	D1*	D2	D3	D4

* Crystals were grown in the well under the condition with 10mg/mL of K78BoF.

△ Crystals were grown in the well under the condition with 20mg/mL of K78BoF.

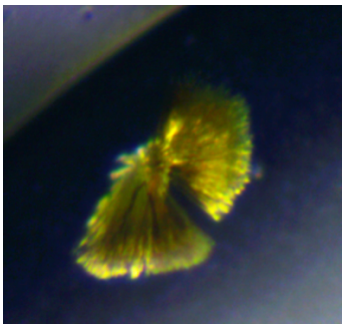
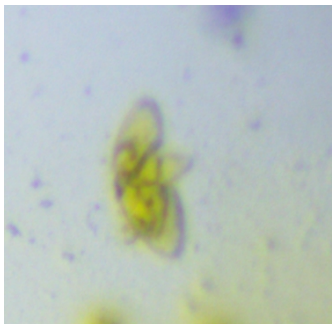
A  B  C 

Fig. 18 Some representative crystals grown in microseeding experiment II with 20mg/mL of K78BoF. (A) Needle-like crystals grown under A4 condition. (B) Crystals grown in a lump under C2 condition. (C) Plate-shape crystals grown under C4 condition.

Despite salt concentration being a small factor compared to pH value of the buffer or concentration of precipitants in crystal growth, it was worthy of a trial since no ideal crystals were obtained in the previous experiments. Microseeding experiment III was designed, as shown on Table 7, which was similar to microseeding experiment II, except for the addition of 0.05M Na_2HPO_4 rather than 0.1M. Unfortunately, no significant improvement in crystal quality or X-ray diffraction results was gained.

Table 7 24-well plate arrangement for microseeding experiment III with 0.05M Na_2HPO_4 and 1000-fold diluted seed stock at 4℃.

pH	4.2	5.0	6.0	7.0	8.0
35%(v/v) PEG 400	A1*△	A2	A3*	A4*△	A5*△
35%(v/v) PEG 500	B1*△	B2*	B3*△	B4	B5
30%(w/v) PEG 1000	C1*△	C2*△	C3*△	C4	C5
20%(w/v) PEG 1000	D1*	D2*	D3	D4	D5

* Crystals were grown in the well under the condition with 10mg/mL of K78BoF.

△ Crystals were grown in the well under the condition with 20mg/mL of K78BoF.

After trying to change several parameters in the condition, the best crystallization condition seemed to be the original one, MCSG-1: G11. The crystal grown in the c2 condition of microseeding experiment I offered a clear diffraction pattern with a high resolution. Its crystal structure was analyzed below.

1.2.3 Optimization of MCSG-1: G10 condition

Increasing protein concentration, changing the ratio of protein solution to crystallization reagent, and microseeding experiments were executed when I tried to optimize MCSG-1: G10 condition for crystallizing K78BoF. Unfortunately, no crystals were found on the plates. Considering the limited amount of the protein we had, further investigation in this part was stopped.

1.3 Analysis on crystal structure of K78BoF

After screening dozens of crystal samples at the X-ray beamline, only one provided clear enough single-crystal monochromatic diffraction patterns with resolution to 2.13Å for 3D-structure determination at pH 6. The monomeric crystal structure (Fig. 19) coincides with characteristics of GFP-like proteins. Eleven connected β-strands construct a cylinder surrounding the chromophore with an α-helix running through the axis of the cylinder and linking them together. K78BoF crystallized in monoclinic space group C121 with unit cell dimensions of a = 84.2Å, b = 34.6Å, c = 88.8Å, $\alpha = \gamma = 90^{\circ}$, $\beta = 110.9^{\circ}$ (Table 8).

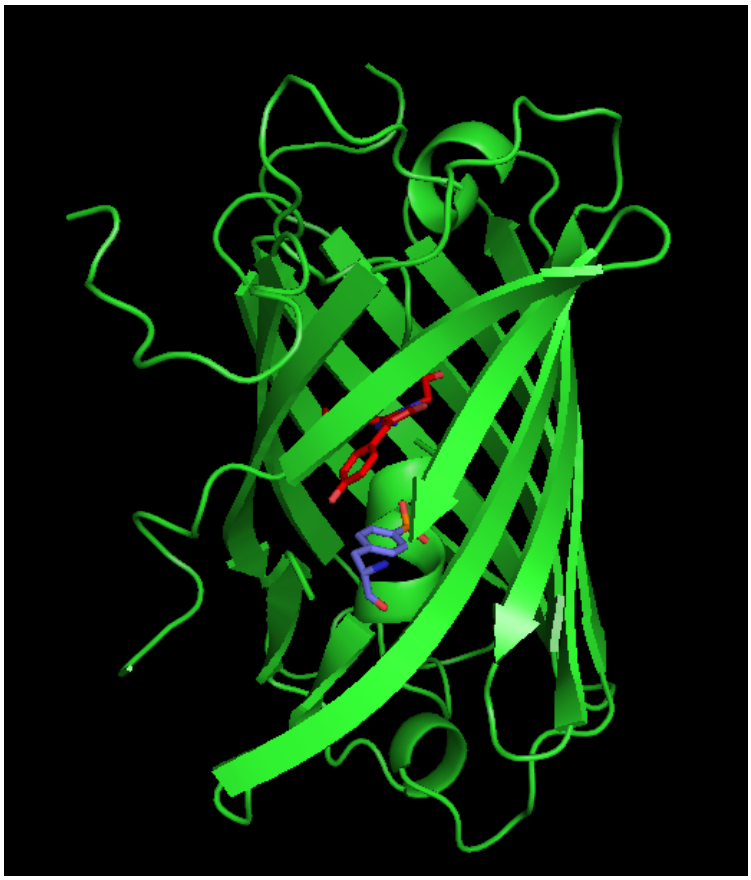


Fig. 19 An overview of crystal structure of protein K78BoF at pH 6. The chromophore is shown in red stick representation while BoF is shown in blue stick representation with carbon atoms in green color, oxygen in red and nitrogen in blue.

Table 8 Crystallographic data for crystal structure of K78BoF.

Space group	C121
Cell dimensions: a (Å)	84.203
Cell dimensions: b (Å)	34.567
Cell dimensions: c (Å)	88.815
Resolution (Å)	83.0-2.13
Completeness (%)	94.89
R/R_{free}	0.1894/0.2608
Rmsd bond lengths (Å)	0.0352
Rmsd bond angles (°)	1.9395

Formed by Met175-Tyr176-Gly177, the chromophore of K78BoF fits well in the electron density map at a sigma level of 1.0 (Fig. 20A) and exists in the *cis* state at pH 6 without any electron density observed in the position for the *trans* state. Two water molecules were located near the chromophore, near Gln173, Gln218, and Ser220 with interatomic distances less than 3.5Å to them (Fig. 21). Hydrogen bonds on the terminal carbonyl oxygen of the conjugated π system are believed to stabilize the excited state and decrease the HOMO-LUMO energy difference, resulting in red-shift of the absorbance [56]. The amino group on the side chain of Arg204 is also close to the chromophore with an interatomic distance of 2.7Å to another carbonyl oxygen, implying a possible hydrogen bond and a potential contribution to the red fluorescence emitted.

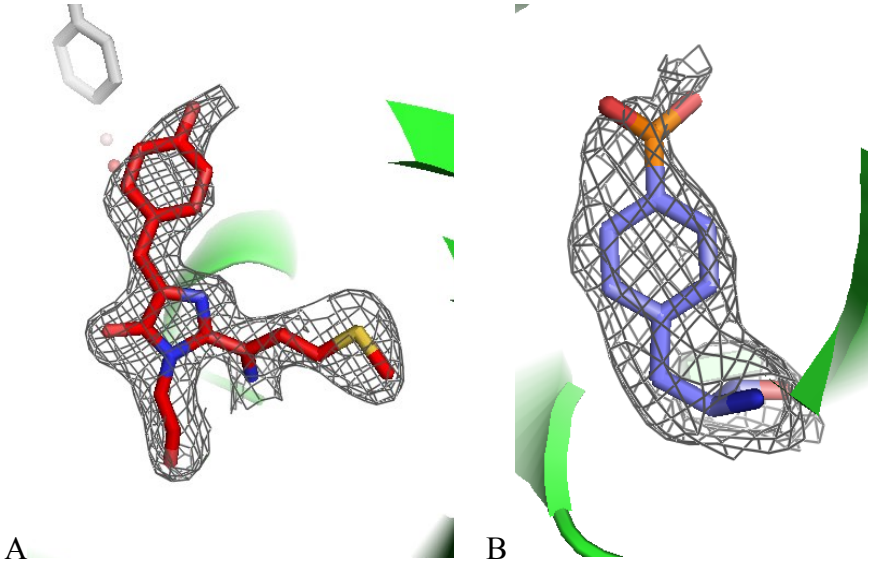


Fig. 20 Structures of K78BoF chromophore (A) and BoF (B) in electron density maps. The chromophore is shown in sticks in an electron density map at 1.0 sigma level while BoF is shown in sticks in a map at 0.5 sigma level.

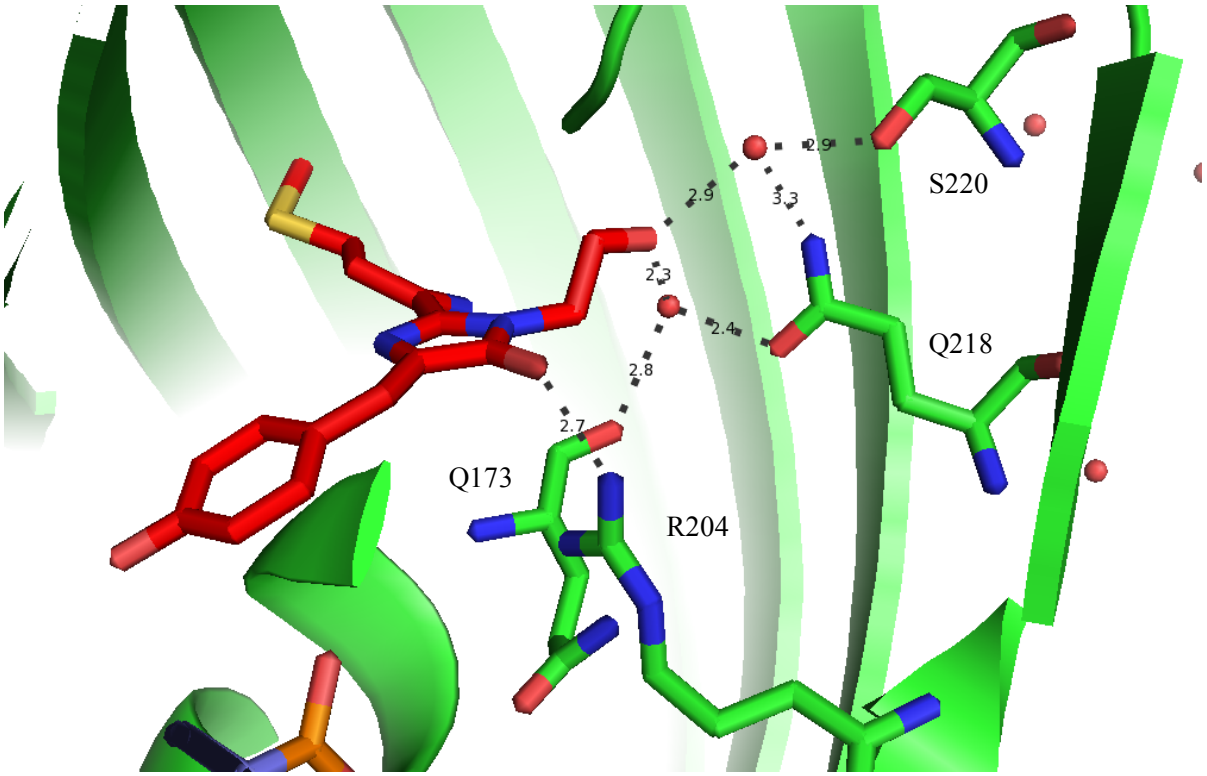


Fig. 21 Possible hydrogen bonds (shown in dotted lines) connecting to the chromophore of K78BoF. Bond lengths were measured by PyMOL with units in Å.

To analyze the effect of the replacement of BoF to Lys30 on the conformation of the protein, the crystal structure of R-GECO1 (PDB 4I2Y) was obtained from Protein Data Bank to compare with K78BoF [72]. R-GECO1 is a tetrameric mApple variant with a difference of four residues in the β-barrel [72]. As can be seen in Fig. 22A, K78BoF is well superimposed upon the backbone of one subunit of R-GECO1, indicating that no significant structural alternation happens because of the substitutions. The five-member rings and the terminal carbonyl groups of chromophores perfectly overlap with each other, yet the phenolic moiety of K78BoF deviates by 7.4º as calculated by PyMOL (Fig. 22B). Moreover, the mutation breaks the hydrogen bonds of the hydroxyl oxygen on the phenolic moiety. In the crystal structure of R-GECO1, a hydrogen bond exists between Lys78 and the chromophore when another hydrogen bond between Ser62 and Lys78 stabilizes it (Fig. 23B). However, after the Lys is replaced by BoF, the interatomic distances are extended to 4.2Å between BoF and the chromophore and 11.5Å between BoF and

Ser14 (Fig. 23A), both of which are too far to build hydrogen bonds. It could be a reason for the angle deviation of the phenolic moiety of K78BoF.

An important point to mention is that the electron density for the two hydroxyl groups can hardly be found bonding to the boron atom on BoF, even at a sigma level of 0.5 (Fig. 20B). Since the *in vivo* stability of BoF has been proven for years [74], this situation is probably caused by other reasons, instead of a chemical change.

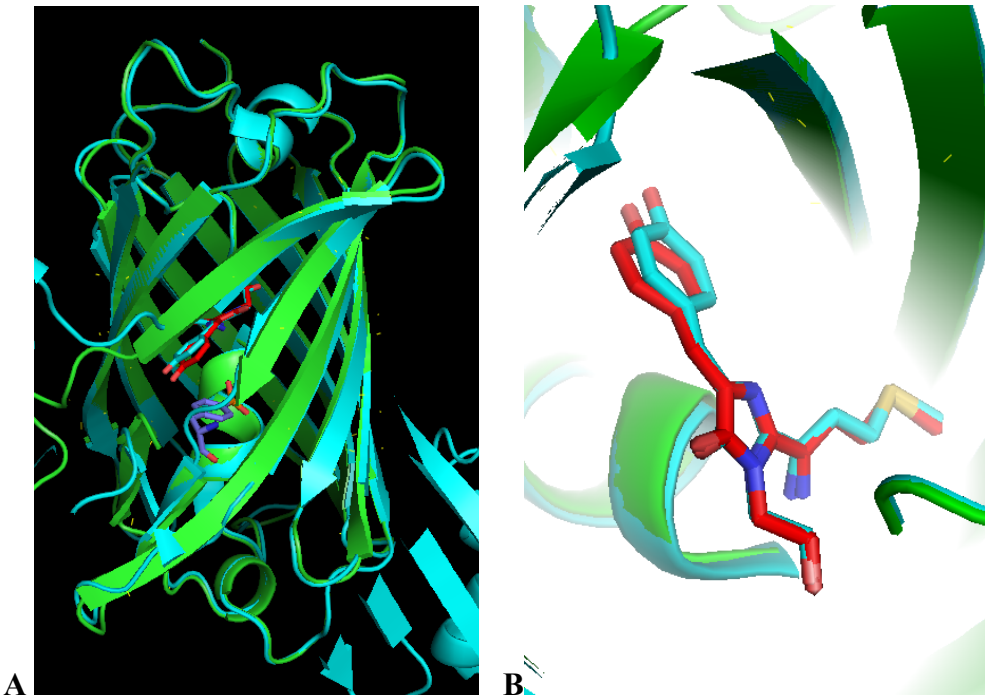


Fig. 22 Alignment of K78BoF crystal structure at pH 6 with R-GECO1. (A) The structure of K78BoF is in green and R-GECO1 is cyan. (B) The chromophore structure of K78BoF is in red color and R-GECO1 is cyan.

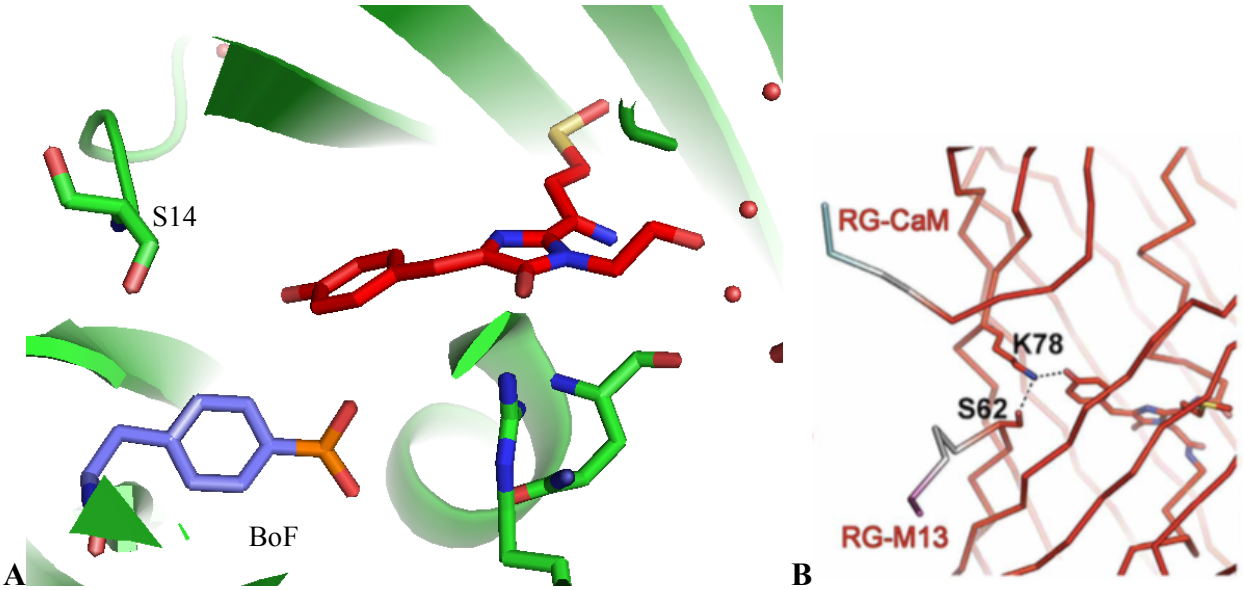
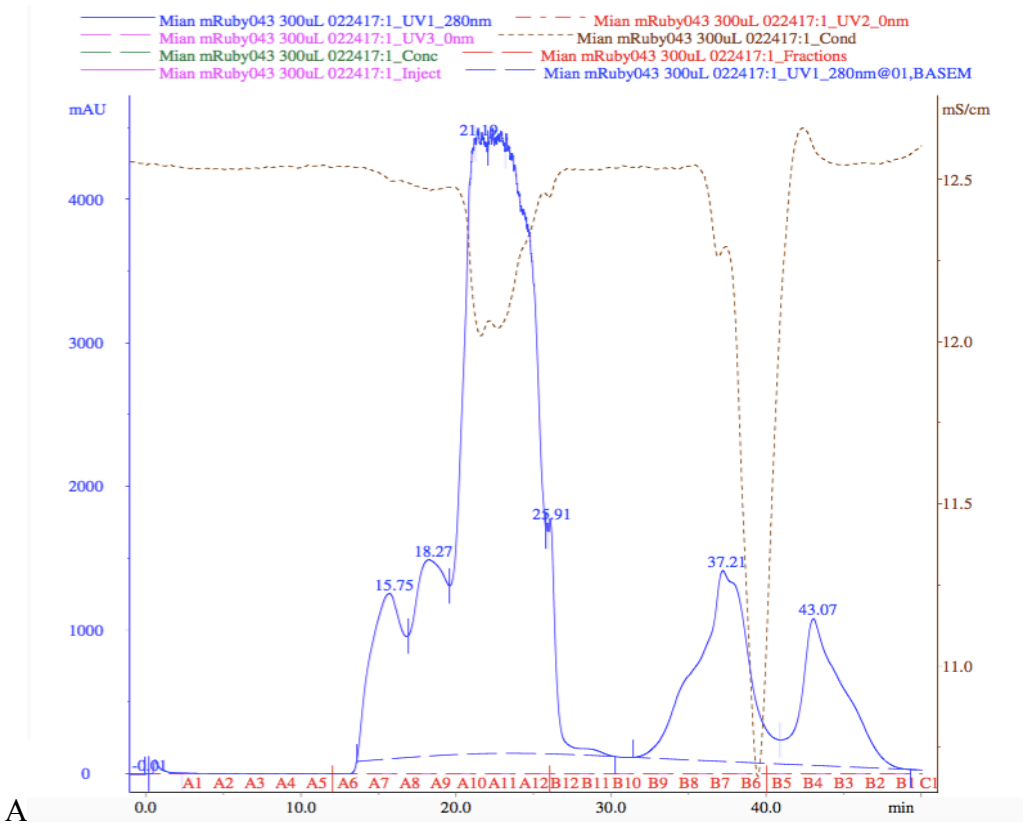


Fig. 23 Distance among chromophores, serine and lysine residues or BoF in K78BoF (A) and R-GECO1 (B) crystal structures [71].

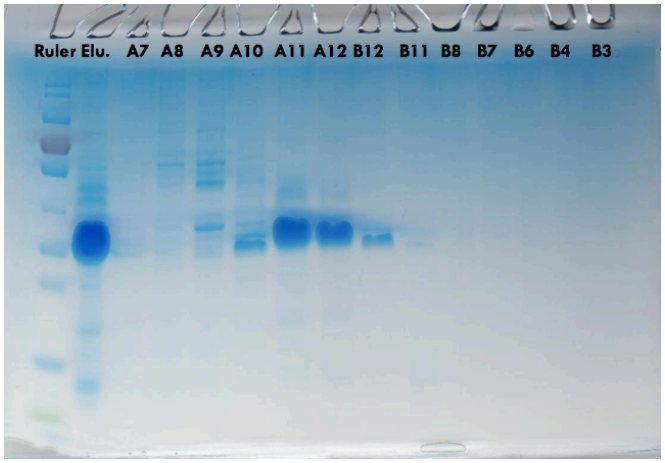
2. Crystallization of mRuby0.4-3

2.1 Expression and purification of mRuby0.4-3

The plasmid bearing the mRuby0.4-3 gene was transformed into *E. Cloni* (Lucigen) cells through heat shocking at 42℃ for 45 seconds after cooling on ice for 30 minutes. The cells were then cultured in TB with ampicillin at 37℃ overnight, followed by decreasing temperature to 18C° overnight to overexpress the protein. The cells were harvested and washed by lysis buffer (50mM Tris, 25mM NaCl, 1mM PMSF, pH 7.5), then lysed by sonication. After centrifuging down the cell fragments, the supernatant was collected via nickel-column purification, followed by size exclusion chromatography (Fig. 24A). As can be seen on the gel (Fig. 24B), fractions A9, A10, A11, A12, and B12 contained mRuby0.4-3. Because there was significant amounts of contaminants in A9, I excluded A9 when combining the fractions to avoid introducing contaminants into the purified protein. Purity of mRuby0.4-3 was analyzed through running a SDS-PAGE gel (Fig. 25). After two-step purification, most of the other proteins were removed. Protein concentration was measured to be about 25.5mg/mL in Nanodrop at 280nm.



A



B

Fig. 24 Size exclusion chromatography of mRuby0.4-3. (A) The protein separation was traced by UV detector. (B) A SDS-PAGE gel was run for fractions containing proteins. On the gel, "Elu" means protein eluent after nickel-column purification.

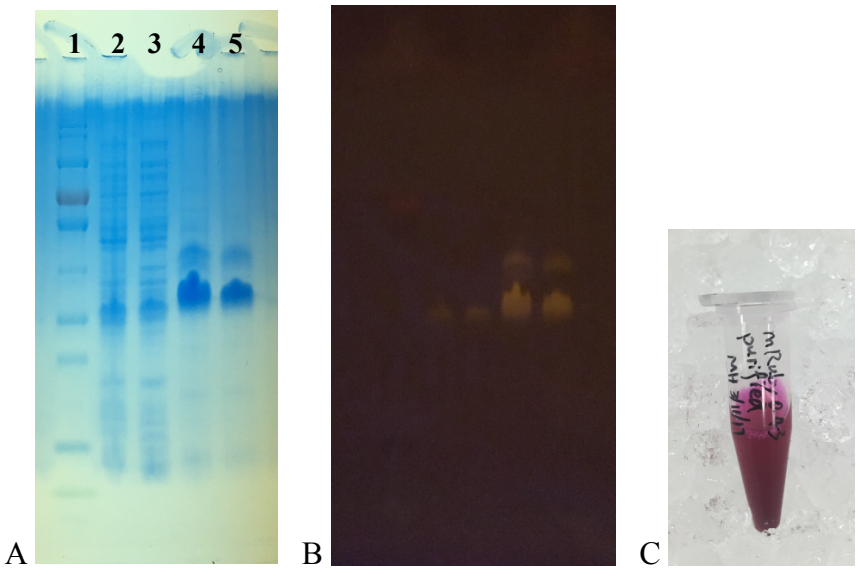


Fig. 25 Purity of mRuby0.4-3 checked by SDS-PAGE gel (A, under visible light; B, under blue light) and the purified protein solution (C). On (A), lane 1 was for prestained protein ladder, lane 2 and 3 for cell pellet and supernatant respectively after lysis, lane 4 for protein solution after nickel-column purification, lane 5 for protein solution after size exclusion.

2.2 Crystallization

2.2.1 Crystallizing method screening

To quickly identify suitable conditions for crystallizing protein mRuby0.4-3, all four MCSG suites were applied. After two weeks of incubation, 10 conditions were found generating crystals in MCSG-1 suite and 4 in MCSG-2 suite. Also, a condition was found growing crystals in MCSG-3 suite after almost four weeks of incubation. Information for these 15 conditions and crystal types grown in them are listed in Table 9.

Table 9 Conditions able to grow crystals for mRuby0.4-3.

Condition	Salt	Buffer	Precipitant	$V_{protein}$: $V_{reagent}$	Crystal Type*
MCSG-1: A2		0.1M CHES: NaOH pH 9.5	30%(w/v) PEG 3000	3:1	3D
MCSG-1: C2	0.2M Lithium Sulfate	0.1M Bis-Tris: HCl pH 5.5	25%(w/v) PEG 3350	3:1	1D
MCSG-1: H3	0.2M Lithium Acetate		20%(w/v) PEG 3350	1:1	3D
MCSG-1: E4	0.2M Lithium Sulfate	0.1M Tris: HCl pH 8.5	25%(w/v) PEG 3350	1:1	3D(s**)
MCSG-1: F5	0.2M Sodium Acetate		20%(w/v) PEG 3350	1:1	3D
MCSG-1: G7		0.1M Tris: HCl pH 8.5	25%(w/v) PEG 3350	1:1	3D
MCSG-1: D8	0.1M Sodium Chloride	0.1M Bis-Tris: HCl pH 6.5	1.5M Ammonium Sulfate	1:3	1D(s)
MCSG-1: D9	0.2M Sodium Chloride	0.1M Tris: HCl pH 8.5	25%(w/v) PEG 3350	1:1	3D(s)
MCSG-1: C11	0.2M Calcium Acetate	0.1M Tris: HCl pH 7.0	20%(w/v) PEG 3000	1:1	3D
MCSG-1: E11			1.0M NaH_2PO_4/ K_2HPO_4 pH 8.2	1:1	2D, 3D
MCSG-2: E3		0.1M Bis-Tris Propane: NaOH pH 7.0	1.5M Lithium Sulfate	1:3	2D
MCSG-2: B4			1.8M NaH_2PO_4/ K_2HPO_4 pH 8.2	3:1	1D(s)
MCSG-2: C7		0.1M Tris: HCl pH 8.5	1.5M Ammonium Phosphate Dibasic	3:1	3D
MCSG-2: F7	0.15M DL-Malic Acid pH 7.0		20%(w/v) PEG 3350	1:3	3D(s)
MCSG-3: D3	0.2M Lithium Sulfate	0.1M Tris: HCl pH 7.0	2.0M Ammonium Sulfate	1:1	3D

* According to the shape of crystals, they usually can be divided into three different types, 1D crystals for spine/needle-like shape, 2D crystals for thin plate shape, 3D crystals for chunk shape.

** The crystals are too small for X-ray diffraction.

Table 10 X-ray diffraction results for selected 3D crystals of mRuby0.4-3.

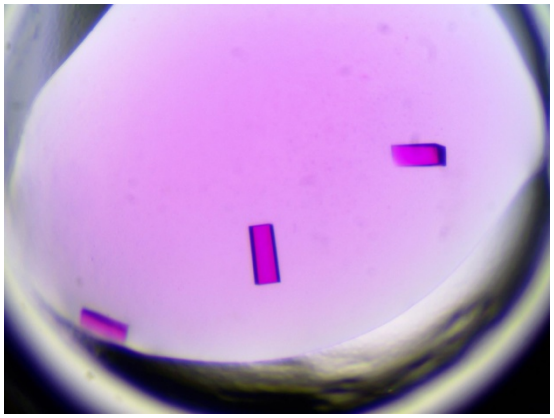
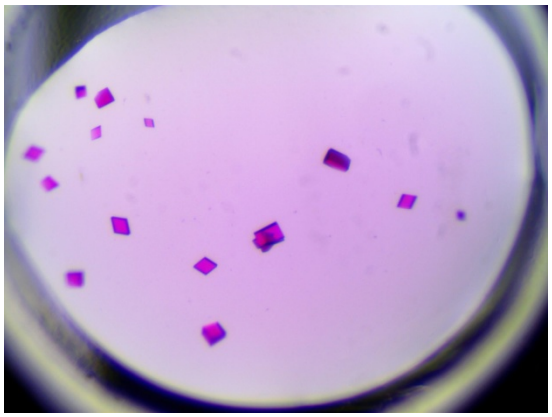
Crystals	X-ray diffraction Results with growth conditions
	**MCSG-1: A2** 1st: Two crystals were prepared but one was lost during cryostorage and the other one gave poor diffraction resolution to 7Å. 2nd: Several crystals were analyzed and one of them showed good resolution at 2.83Å with clean diffraction pattern.
	**MCSG-1: F5** 1st: One piece of crystals was analyzed but no data was obtained because the diffraction was highly smeared.

Table 10 (Continued) X-ray diffraction results for selected 3D crystals of mRuby0.4-3.

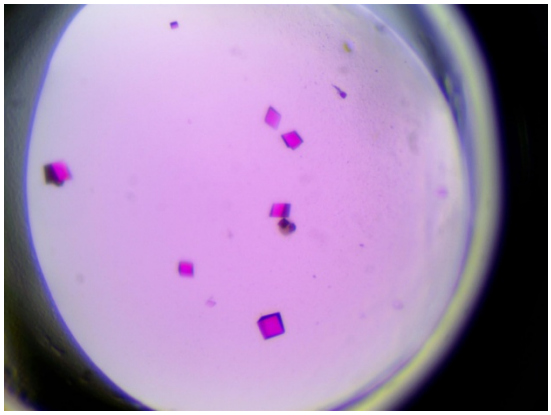
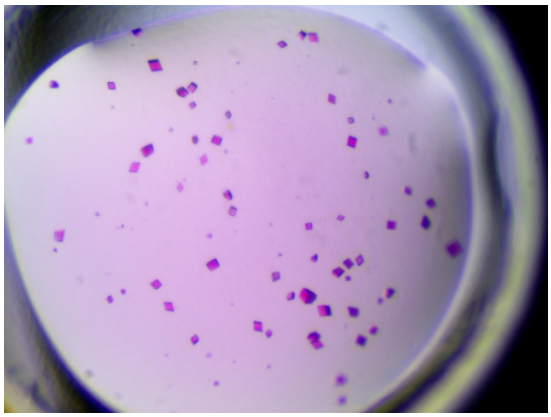
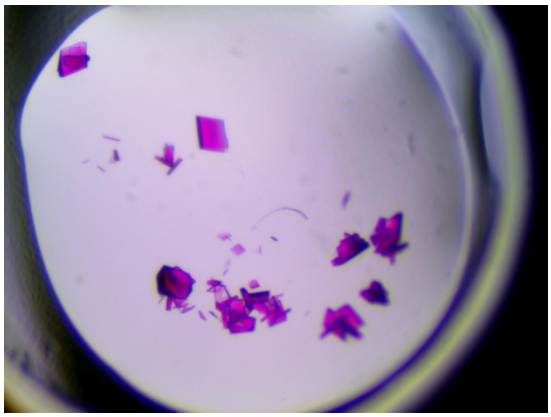
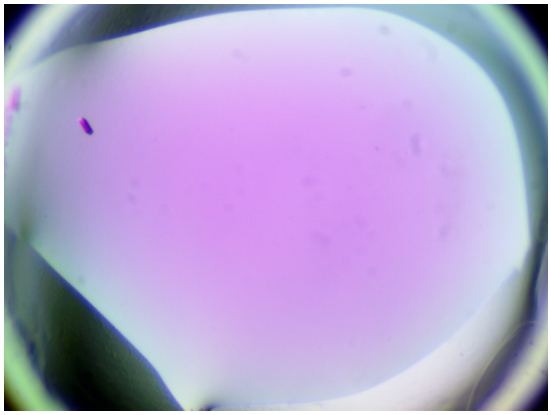
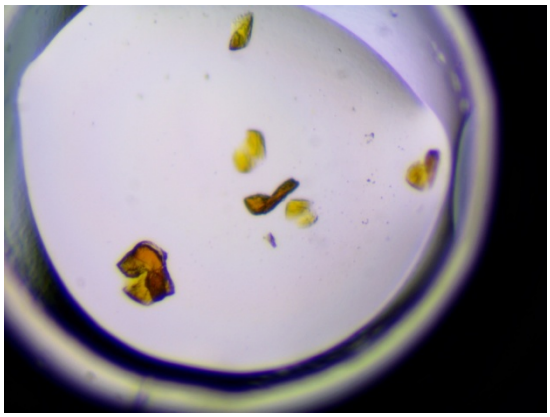
Crystals	X-ray diffraction Results with growth conditions
	**MCSG-1: G7** 1st: Poor resolution of a piece of crystal at 12Å was observed from X-ray diffraction.
	**MCSG-1: C11** 1st: Two pieces of crystals were picked for X-ray diffraction. One was lost during cryostorage. The other one was successfully tested with a resolution at 3.0Å but images were smeared. 2nd: Five crystals were tested, but all of them showed smeared diffraction.
	**MCSG-1: E11** 1st: Poor resolution at 8Å with smeared diffraction with a piece of crystal in this condition. 2nd: Three crystals were tested. One of them displayed a clean diffraction pattern with 2.6Å resolution. The other two also had good resolution, but the crystals were cracked.

Table 10 (Continued) X-ray diffraction results for selected 3D crystals of mRuby0.4-3.

Crystals	X-ray diffraction Results with growth conditions
	**MCSG-2: C7** 1st: Same as the one in condition MCSG-1: E11, the crystal displayed poor resolution at 8Å with smeared data.
	**MCSG-3: D3** 2nd: Four crystals were sent for analysis, but all of them had diffraction that were highly smeared.

Usually, 3D crystals with clear edges give better resolution and diffraction patterns in X-ray analysis. Crystals grown in conditions MCSG-1: A2, F5, G7, C11, E11, MCSG-2: C7, and MCSG-3: D3 were prepared with cryo-protectant LVCO and sent for the first X-ray shooting. The pictures of the crystals and the diffraction results were described on Table 10. The diffraction for most of crystals from different conditions were smeared, demonstrating that LVCO may be not a suitable cryo-protectant for them.

Given the many shapes and numbers of the crystals, I selected the ones from MCSG-1: A2, MCSG-1: C11, MCSG-1: E11, and MCSG-3: D3 for the second X-ray shooting. Instead of LVCO, they were prepared with cryo-protectants made of crystallization reagents with 10%

ethylene glycol (for MCSG-1: A2 and MCSG-1: C11) or 25% glycerol (for MCSG-1: E11 and MCSG-3: D3). The results were also displayed in Table 10. Crystals from MCSG-1: A2 and MCSG-1: E11 with clean diffraction patterns and high resolution were processed for the structure model building steps. However, both of these two crystals were produced in a basic environment. Since the goal of this project was to investigate the change of the protein conformation with the pH value, crystals grown in acidic and neutral conditions are still required.

2.2.2 Optimization of crystallizing method

In this part, I focused on optimizing the crystallization condition MCSG-1: A2, MCSG-1: C2, MCSG-1: C11, MCSG-1: E11, and MCSG-3: D3 though changing precipitant concentration and/or protein concentration, and using the microseeding method.

Although crystals from MCSG-1: A2 provided acceptable diffraction to solve the 3D structure, further optimization of crystallization method was still conducted to look for crystals with higher resolution diffraction. Because our lab does not have CHES, the buffer contained in the condition, I could not prepare the crystallization reagent. Yet I still tried to increase the protein concentration (Table 11) and the proportion of the protein in the volume ratio, and added seed solution into the system (Table 12). However, better crystals were not found by X-ray shooting.

MCSG-1: C2 was the most acidic condition found being able to grow crystals, even though they were needle-like (Fig. 26). Optimization of this condition might assist in obtaining a protein structure in an acidic environment. Results from the experiments (Table 12 and 13) showed that microseeding favored the crystals growing in lower protein or precipitant concentration, but the crystals still grew as needles too thin for X-ray diffraction.

MCSG-1: C11 could provide a condition with pH at 7.0 for crystal growth, so it was worth doing more work on it, but good results were not obtained from X-ray diffraction, even for crystals grown with good shape (Table 11).

Table 11 Design of optimization experiment I for condition MCSG-1: C2, MCSG-1: C11, MCSG-1: E11 and MCSG-3: D3.

Condition△	Protein concentration (mg/mL)			
	13	26	13	26
MCSG-1: A2*	30%(w/v) PEG3000 C	30%(w/v) PEG3000 C	—	—
MCSG-1: C2*	25%(w/v) PEG3350	25%(w/v) PEG3350	20%(w/v) PEG3350	20%(w/v) PEG3350
MCSG-1: C11**	20%(w/v) PEG3350 C	20%(w/v) PEG3350 C	15%(w/v) PEG3350	15%(w/v) PEG3350
MCSG-1: E11**	1.0M NaH_2PO_4/ K_2HPO_4 pH 8.2 C	1.0M NaH_2PO_4/ K_2HPO_4 pH 8.2 C	0.75M NaH_2PO_4/ K_2HPO_4 pH 8.2	0.75M NaH_2PO_4/ K_2HPO_4 pH 8.2
MCSG-3: D3**	2.0M Ammonium Sulfate	2.0M Ammonium Sulfate	1.5M Ammonium Sulfate	1.5M Ammonium Sulfate

△ For each condition, only concentration of precipitant was altered, salt and buffer was same to the original.

* $V_{protein}$: $V_{reagent}$ = 3:1.

** $V_{protein}$: $V_{reagent}$ = 1:1.

C Crystals were grown in this condition.

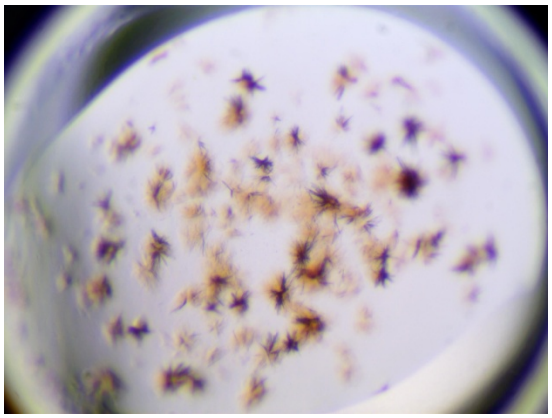


Fig. 26 Needle-like crystals grown in condition MCSG-1: C2.

Table 12 Microseeding experiment for condition MCSG-1: A2, MCSG-1: C2, MCSG-1: E11 and MCSG-3: D3.

Condition[△]			
MCSG-1: A2	MCSG-1: C2*	MCSG-1: E11**	MCSG-3: D3**
$V_{protein}$: $V_{reagent}$: V_{seed} = 6:1:1 [C]	25%(w/v) PEG3350 [C]	1.0M NaH_2PO_4/ K_2HPO_4 pH 8.2 [C]	2.0M Ammonium Sulfate
$V_{protein}$: $V_{reagent}$: V_{seed} = 3:2:1	20%(w/v) PEG3350 [C]	0.75M NaH_2PO_4/ K_2HPO_4 pH 8.2	2.0M Ammonium Sulfate

[△] For each condition except MCSG-1: A2, only concentration of precipitant was altered, salt and buffer was same to the original.

* $V_{protein}$: $V_{reagent}$: V_{seed} = 6:1:1.

** $V_{protein}$: $V_{reagent}$: V_{seed} = 3:2:1.

[C] Crystals were grown in this condition.

Table 13 Design of optimization experiment II combining with microseeding method* for condition MCSG-1: C2.

MCSG-1: C2[△]	Protein concentration (mg/mL)	
	6.5	13
15%(w/v) PEG3350	A1	A2
20%(w/v) PEG3350	B1	B2 [C]
25%(w/v) PEG3350	C1 [C]	C2 [C]

[△] For each condition, only concentration of precipitant was altered, salt and buffer was same to the original.

* $V_{protein}$: $V_{reagent}$: V_{seed} = 6:1:1.

[C] Crystals were grown in this condition.

Similar to MCSG-1: A2, although a clean diffraction pattern was received, optimization of MCSG-1: E11 was still performed to obtain crystals that diffract to higher resolution. According to the results on Table 11 and 12, crystals could not be generated in lower precipitant concentrations, even though the microseeding method was applied. Some crystals were selected and ready for analysis by X-ray.

The brownish color of the crystals grown in MCSG-3: D3 (Table 4) deviated from the original color of the protein, which was interesting to note for further investigation. However, the crystals could not be reproduced, even by microseeding method.

2.3 Analysis of crystal structure of mRuby0.4-3

mRuby0.4-3 at pH 8.2 provided clear diffraction patterns with resolution at 2.63Å from X-ray diffraction experiment, so a dimeric structure model (Fig. 27) including two completely identical chains, chain A and chain B, was built based on the data. Similar to K78BoF, it showed typical characteristics of GFP-like proteins. The protein also crystallized in monoclinic space group C121 but with different unit cell dimensions: a = 108.4Å, b = 77.4Å, c = 66.9Å, $\alpha = \gamma = 90^{o}$, $\beta = 92.1^{o}$ (Table 14).

Although the chromophore was the same as K78BoF, according to electron density maps, mRuby0.4-3 chromophores are in the *trans* state (Fig. 28), rather than the *cis* state. Meanwhile, several potential hydrogen bonds were found between the chromophore and Thr64, Arg71, Arg96 with interatomic distances no more than 3.5Å, probably to assist with stabilization of the chromophore and red-shift the absorbance. Given that the resolution of the crystal structure was not high enough to recognize water molecules, hydrogen bonds between water molecules and chromophores, which usually play important roles in protein excitation and red-shift cannot be studied here. However, some hints were still able to be seen on the Fo-Fc map. The green electron density shown on Fig. 30 suggest that certain atoms exist at those sites, which could either be water molecules or free ions from crystallization reagents. To prove that, a crystal model with higher resolution will be required.

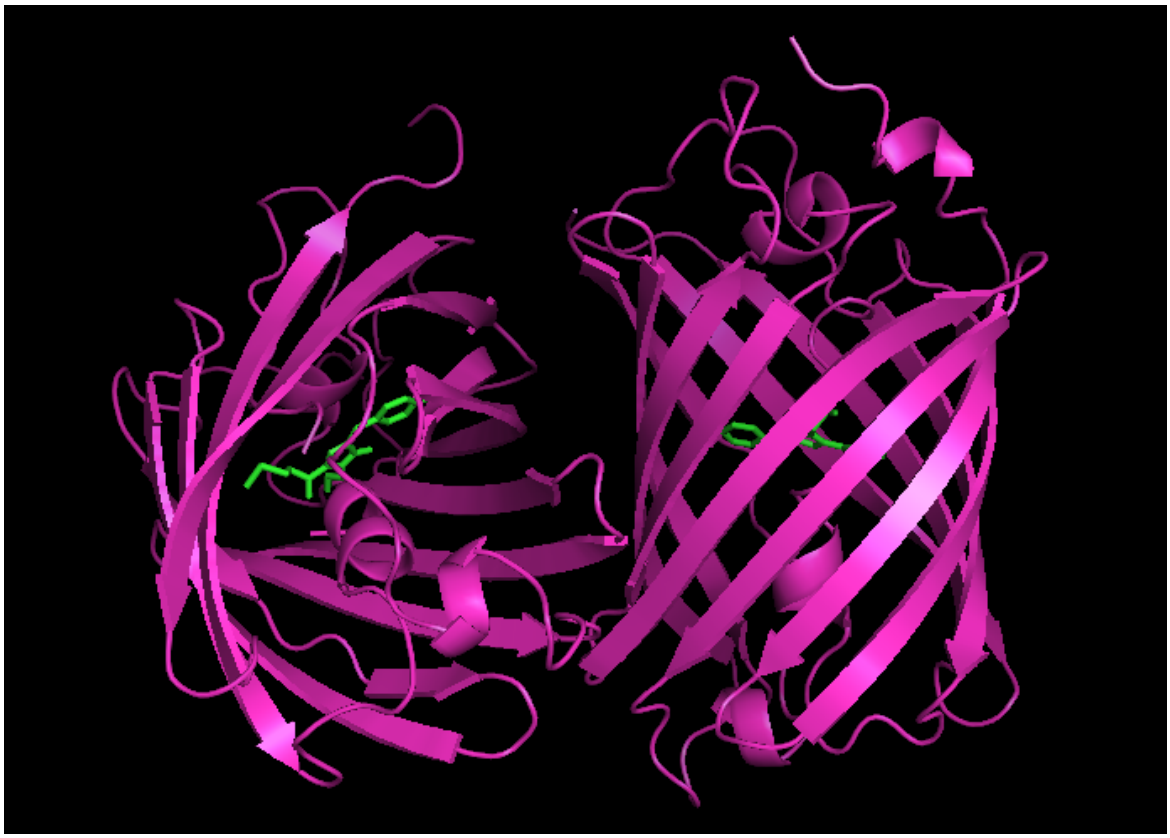


Fig. 27 An overview of crystal structure of mRuby0.4-3 at pH 8.2.

Table 14 Crystallographic data for crystal structure of mRuby0.4-3.

Space group	C121
Cell dimensions: b (Å)	108.446
Cell dimensions: b (Å)	77.390
Cell dimensions: c (Å)	66.851
Resolution (Å)	66.8-2.63
Completeness (%)	99.53
R/R_{free}	0.2276/0.3088
Rmsd bond lengths (Å)	0.0115
Rmsd bond angles (°)	1.8593

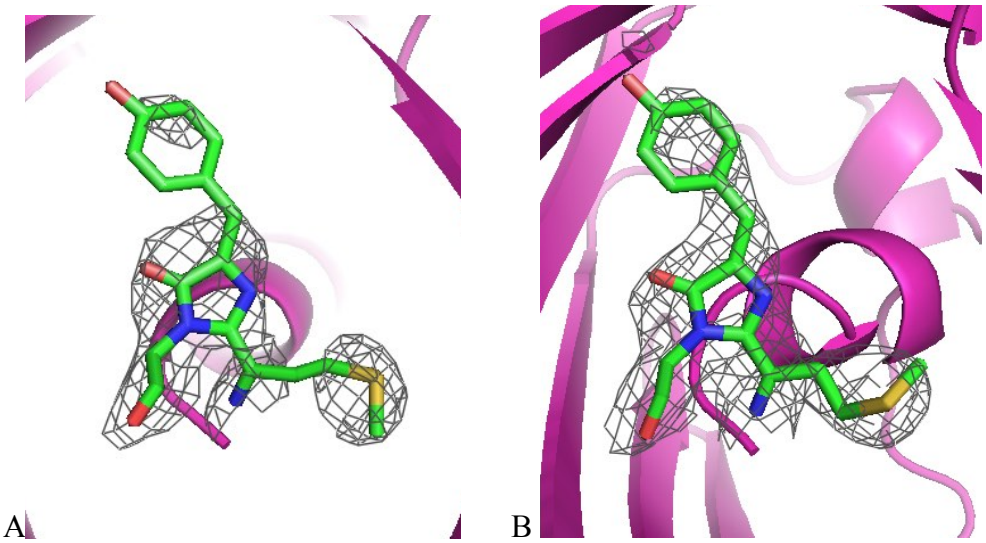


Fig. 28 Chromophore structures of chain A (B) and chain B (B) for mRuby0.4-3 in electron density maps at 1.0 sigma level.

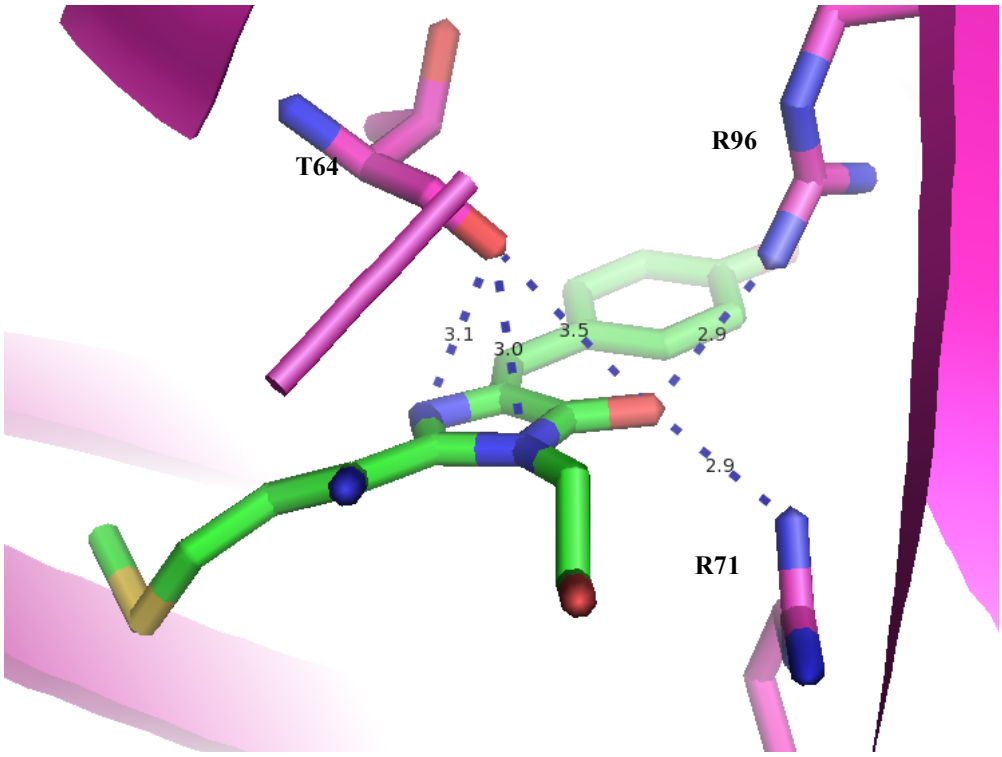


Fig. 29 Possible hydrogen bonds (shown in dotted lines) connecting to the chromophore of mRuby0.4-3. Bond lengths were measured by PyMOL with unit in Å.

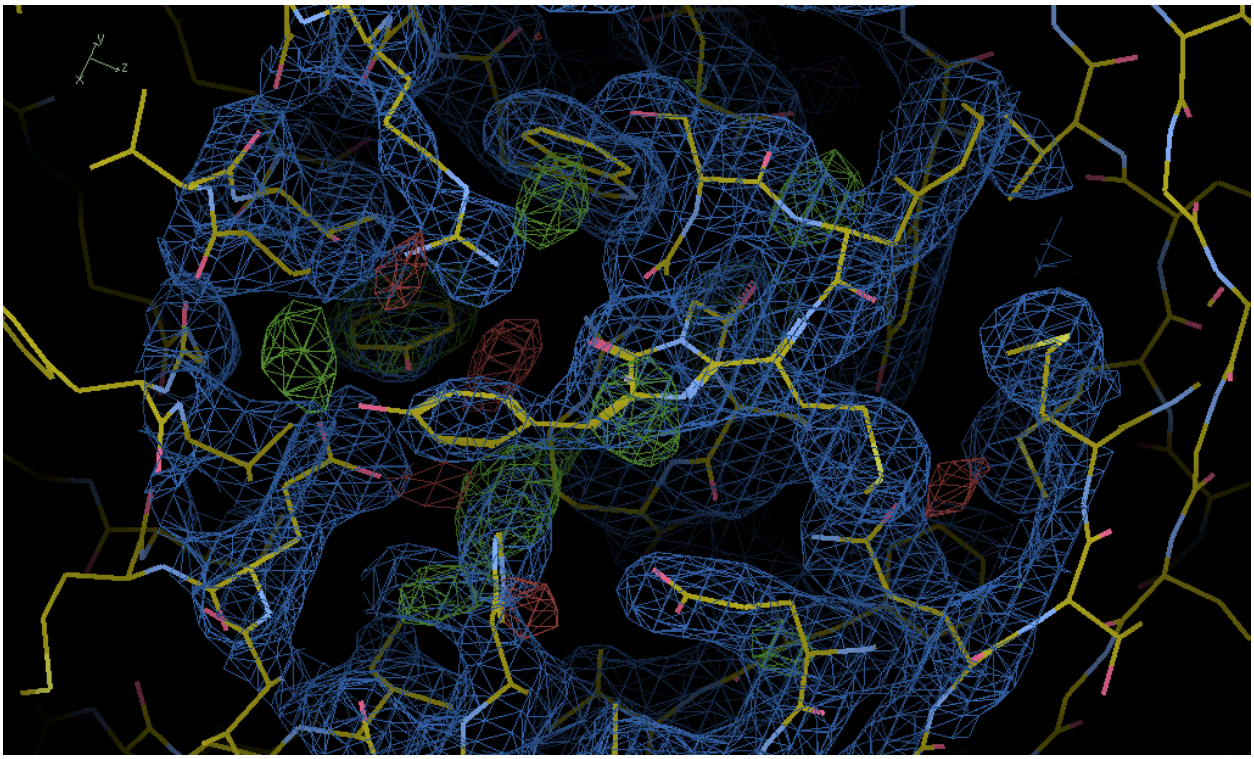


Fig. 30 An overview of residues surrounding the mRuby0.4-3 chromophore in electron density maps. The figure is a screenshot from COOT with 2Fo-Fc map in blue color and Fo-Fc map in green and red.

DISCUSSION

1. Optimization of crystallization methods to improve crystal quality

To obtain good diffraction patterns from X-ray experiments, a large crystal visible under the microscope and a regular shape with certain thickness should be used. In both projects, I tried to improve the quality of the crystals through manipulating crystallization conditions. This included changing incubating temperature, protein concentration and volume, buffer concentration and pH, and salt and precipitant concentration. However, no significant improvement was observed visually, and this was further confirmed later by X-ray diffraction results.

I also introduced the microseeding method to the optimization experiments, which was more effective for K78BoF crystals than mRuby0.4-3 crystals. Moderate improvements on the transparency and brightness of K78BoF crystals were observed, though the sizes and the shapes did not show significant differences. The resolution of one K78BoF crystal improved from 3.0Å to 2.1Å.

Actually, there is a chance to further improve the resolution of K78BoF crystals by applying the serial seeding method. Serial seeding is an advanced microseeding method that repeatedly seeds a new drop made by crystals from the last run of microseeding experiments. It can usually help to improve the resolution of crystals in a better manner [75]. For example, the resolution of ErbB-3-binding protein 1 crystal was increased from 3.2Å to 1.6Å after two runs of serial seeding [76]. Although wet-lab work for the K78BoF project is finished, this method may be employed to improve the quality of mRuby0.4-3 crystals.

2. BoF in crystal structures

In spite of boronic acids not being found in nature and the need to be prepared from primary sources of boron, they have been proven stable *in vivo*, nontoxic to cells and are now applied commonly in biomedical field studies for developing receptors and sensors for carbohydrates and

some small molecules, antimicrobial agents, enzyme inhibitors, and anti-cancer drugs [77]. However, conjugating it with FPs to develop FP-based biosensors is still a new research field. In 2012, Wang et al substituted Tyr66 at the chromophore of a GFP with BoF to form a mutant UFP-Tyr66pBoPhe with a crystal structure solved at a resolution of 1.24Å, reporting its H_2O_2-sensing ability in *E. coli* [78]. In 2013, the Ai group reported the first genetically encoded fluorescent probe for peroxynitrite, pnGFP, in which BoF replaced Tyr174 at the chromophore[63]. They further mutated pnGFP into pnGFP1.5 and tried to determine their crystal structures, but no crystals were grown [79]. In 2016, they developed another peroxynitrite biosensor, K78BoF. Unlike pnGFP, K78BoF was derived from RFP rather than GFP, and its BoF mutagenesis happened on the β-barrel, instead of at the chromophore. Using the purified protein provided by the Ai group, we successfully crystallized K78BoF at pH 6.

The electron density map perfectly covers the boron atom and the phenylalanine moiety of BoF, but the two hydroxyl groups are excluded by the map, even at 0.5 sigma level (Fig. 20B). One explanation for this phenomenon is that inconsistent conformation of BoF existed in the protein crystal. In the crystal structure of phenylboronic acid, the weak B-C π-bonding and the intermolecular hydrogen bonds stabilized the conformation and made the CBO_2 plane coplanar to the benzene ring; but asymmetric forces on the two hydroxyl groups might cause a twist between two planes [77]. In the crystal structure of UFP-Tyr66pBoPhe, the CBO_2 plane was rotated slightly owing to two hydrogen bonds formed with the nitrogen of His148 and the hydroxyl group of Ser205 [78]. Checking adjacent area of BoF in K78BoF structure model, I could not find any residue or water molecule in a reasonable distance ($\leq$ 3.5Å) to form a hydrogen bond. Without an outer restriction, the CBO_2 planes probably occurred with different angels to the benzene ring in different K78BoF molecules. Given that model building is based on all the diffraction patterns generated from all molecules shot by X-rays, the inconsistent position of two hydroxyl groups on the CBO_2 plane may lead to weak X-ray scattering and poor electron density.

3. The missing loop in the K78BoF crystal structure

A loop from Gly100 to Met114 was found missing during model building after molecular replacement. During model refinement, I tried to add the residues back to the model manually, but then realized that they could hardly be fitted into the electron density map because the electron distribution in the area expected for the loop was weak and diffuse. Also, their inclusion in the model increased R factor and R-free factor dramatically. The R factor and R-free factor are indicators for the agreement of the experimental crystal data to a structure model, and the lower the factors are, the better they agree with each other. The addition of the missing loop to the model would reduce the reliability of the crystal structure. Actually, the homologous protein R-GECO1 also missed a loop in a similar position in its crystal structure [71]. The reason for the incomplete model may be complicated, but some clues can still be found.

The Debye–Waller factor, usually known as the B factor or temperature factor, is an indicator for the relative vibrational motion of atoms in a structure. Well-ordered atoms in a structure show low B factor, while it increases with enhanced mobility. Testing B factor for the model by PyMOL, I found very high B factor for the residues on the terminals of the missing loop, as can be seen on Fig. 31. It implies that the missing loop could be so flexible in the structure that the electron density cannot be mapped. The result is not surprising if we check the components of the loop. The majority of the loop is constructed by glycine, serine and lysine, all of which tend to appear frequently in the missing polypeptides of other protein crystal structures reported in the Protein Data Bank [80]. Easy rotation of their side chains in space is believed to increase mobility of the loop, resulting in disorder of the electron distribution on the map.

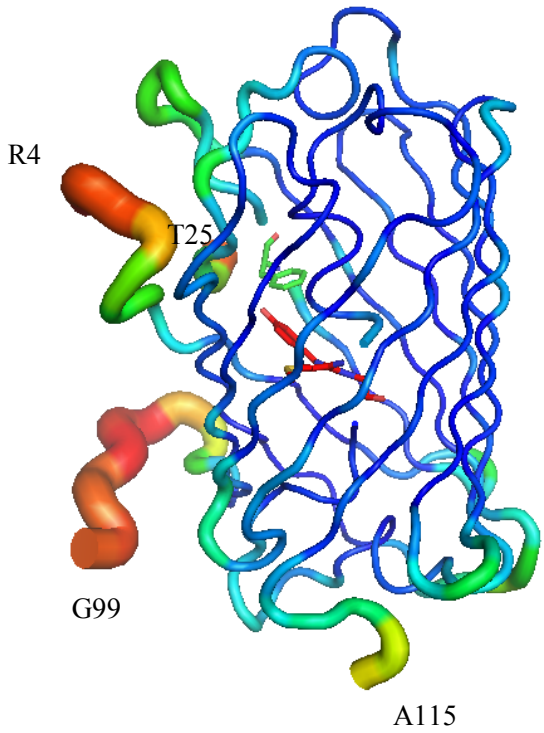


Fig. 31 An overview of K78BoF crystal structure in a B factor version. Red color represents high B factor while blue means B factor for the residues is low.

4. Future work for mRuby0.4-3 project

Until now, I have obtained two mRuby0.4-3 crystals at pH 8.2 and 9.5 respectively and the structure model for the one at pH 8.2 has been built. Reviewing the two goals expected in the mRuby0.4-3 project, the first goal on determining its crystal structure was achieved. Yet for the second goal, a lot of further work needs to be done in future experiments to analyze the effect of pH change on the conformations of the protein.

The crystal grown at pH 9.5 was ready for structure model building. However, crystals grown in neutral and acidic conditions are still needed. Although the needle-like crystals I grew are still far from ideal for X-ray diffraction, the only effective acidic method found in the preliminary screening experiment, condition MCSG-1: C2, is a good starting point for further optimization. Similar to what we have done to crystallize K78BoF, changing diverse factors of

the method in conjugation with the microseeding method may be useful here. The serial seeding method is worthy of a trial as well. A situation that could not be understood was the crystals grown in condition MCSG-1: C11. Even though the crystals show a perfect shape and size for X-ray diffraction, no diffraction was observed from shooting. Considering the neutral condition obtained for growth of the crystal, we still expect to obtain good diffraction data from this method. We will try to conduct serial seeding experiments to look for a breakthrough for this crystallization method.

CONCLUSION

In this thesis, X-ray crystallography of two red fluorescent protein variants, K78BoF and mRuby0.4-3, were studied to decipher their 3D structures. By using purified protein provided by the Ai lab, a monomeric crystal of K78BoF was successfully grown at pH 6 that provided resolution of 2.13Å. The structure model showed that its chromophore is formed by Met-Tyr-Gly in the *cis* state. Two water molecules and several residues on the β-barrel interact with the chromophore through hydrogen bonds which stabilize the conformation. However, comparing it with a homologous variant R-GECO1, we find that the substitution of the unnatural amino acid *p*-boronophenylalanine to Lys prevents the formation of a hydrogen bond between the chromophore and Lys, resulting in a conformation change of the phenolic moiety in the chromophore. This could be a potential reason for the pH-sensitivity of K78BoF.

Protein mRuby0.4-3 was provided as a gene in a plasmid by the Chu lab, so I performed protein expression and purification before it was ready for crystallization. Unlike K78BoF, crystallization of mRuby0.4-3 showed a strong tolerance to pH change and two dimeric crystals grown at pH 8.2 and 9.5 diffracted in X-ray experiments with resolutions of 2.63Å and 2.84Å respectively. In the structure model for the crystal at pH 8.2, the chromophore has a *trans* isomer to K78BoF, with several hydrogen bonds linking the residues surrounding it. However, the resolution of the crystal was not good enough to recognize water molecules in the electron density map, preventing us from deeper structural analysis.

REFERENCES

1. Shimomura O, Johnson FH, Saiga Y. Extraction, purification and properties of Aequorin, a bioluminescent protein from the luminous *Hydromedusan, Aequorea*. *J. Cell. Comp. Physiol*. **59**, 223-39 (1962).

2. Johnson FH, Shimomura O, Saiga Y, Gershman LC, Reynolds GT, Waters JR. Quantum efficiency of *Cypridina* luminescence, with a note on that of *Aequorea*. *J. Cell. Comp. Physiol*. **60**, 85-103 (1962).

3. Shimomura O. Structure of the chromophore of *Aequorea* green fluorescent protein. *FEBS Lett*. **104**, 220-2 (1979).

4. Morise H, Shimomura O, Johnson FH, Winant J. Intermolecular energy transfer in the bioluminescent system of *Aequorea*. 1974. *Biochemistry* **13**, 2656-62 (1974).

5. Perozzo MA, Ward KB, Thompson RB, Ward WW. X-ray diffraction and time-resolved fluorescence analyses of *Aequorea* green fluorescent protein crystals. *J. Biol. Chem*. **263**, 7713-16 (1988).

6. Ormö M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**,1392-95 (1996).

7. Yang F, Moss LG, Phillips GN Jr. The molecular structure of green fluorescent protein. *Nat. Biotechnol*. **14**, 1246-51 (1996).

8. Cody CW, Prasher DC, Westler WM, Prendergast FG, Ward WW. Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. *Biochemistry* **32**, 1212-18 (1993).

9. Herm R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12501-04 (1994).

10. Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. Green fluorescent protein as a marker for gene expression. *Science* **263**, 802-4 (1994).

11. Lippincott-Schwartz J, Patterson GH. Development and use of fluorescent protein markers in living cells. *Science* **300**, 87 (2003).

12. Chalfie M & Kain SR. Green Fluorescent protein – Properties, applications, and protocols. 1-132 (2005).

13. Chudakov DM, Lukyanov S, Lukyanov KA. Fluorescent proteins as a toolkit for *in vivo* imaging. *TRENDS Biotechnol*. **23**, 605-13 (2005).

14. Hunt ME, Scherrer MP, Ferrari FD, Matz MV. Very bright green uorescent proteins from the Pontellid copepod *Pontella mimocerami*. *PLoS One.* **5**, e11517 (2010).

15. Deheyn DD, Kubokawa K, McCarthy JK, Murakami A, Porrachia M, Rouse GW, Holland ND. Endogenous green fluorescent protein (GFP) in amphioxus. *Biol. Bull.* **213**, 95-100 (2007).

16. Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, Matz MV. Diversity of evolution of the green fluorescent protein family. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4256-61 (2002).

17. Brejc K, Sixma TK, Kitts PA, Kain SR, Tsien RY, Ormo M, Remington SJ. Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2306-11 (1997).

18. Palm GJ, Zdanov A, Gaitanaris GA, Stauber R, Pavlakis GN, Wlodawer A. The structural basis for spectral variations in green fluorescent protein. *Nat. Struct. Biol*. **4**, 361-65 (1997).

19. Gurskaya NG, Savitsky AP, Yanushevich YG, Lukyanov SA, Lukyanov KA. Color transitions in coral's fluorescent proteins by site-directed mutagenesis. *BMC Biochem*. **2**, 6 (2001).

20. Lukyanov KA, Fradkov AF, Gurskaya NG, Matz MV, Labas YA, Savitsky AP, Markelov ML, Zaraisky AG, Zhao X, Fang Y. Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *J. Biol. Chem*. **275**, 25879-82 (2000).

21. Wiedenmann J, Elke C, Spindler KD, Funke W. Cracks in the β-can: fluorescent proteins from *Anemonia sulcata* (Anthozoa, Actinaria). *Proc. Natl. Acad. Sci. U.S.A*. **97**, 14091-6 (2000).

22. Gurskaya NG, Fradkov AF, Terskikh A, Matz MV, Labas YA, Martynov VI, Yanushevich YG, Lukyanov KA, Lukyanov SA. GFP-like chromoproteins as a source of far-red fluorescent proteins. *FEBS Lett*. **507**, 16-20 (2000).

23. Yue J, Holland ND, Holland LZ, Deheyn DD. The evolution of genes encoding for green fluoresecent proteins: insights from cephalochordates (amphioxus). *Sci. Rep*. **6**, 28350 (2016).

24. Malo GD, Wang M, Wu D, Stelling AL, Tonge PJ, Wachter RM. Crystal structure and raman studies of dsFP483, a cyan fluorescent protein from *Discosoma striata*. *J. Mol. Biol*. **378**, 871-886 (2008).

25. Henderson JN, S. James Remington, SJ. Crystal structures and mutational analysis of amFP486, a cyan fluorescent protein from *Anemonia majano*. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12712-17 (2005).

26. Remington SJ, Wachter RM, Yarbrough DK, Branchaud B, Anderson DC, Kallio K, Lukyanov KA. zFP538, a yellow-fluorescent protein from Zoanthus, contains a novel three-ring chromophore. *Biochemistry* 44, 202-12 (2005).

27. Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, Lukyanov SA. Fluorescent proteins from nonbioluminescent Anthozoa species. *Nat. Biotechnol*. **17**, 969-73 (1993).

28. Wall MA, Socolich M. Ranganathan R. The structural basis for red fluorescence in the tetrameric GFP homolog DsRed. *Nat. Struct. Biol*. **7**, 1133-8 (2000).

29. Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY. The structure of the chromophore within DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11990-5 (2000).

30. Wiedenmann J, Schenk A, Rocker C, Girod A, Spindler KD, Nienhaus GU. A far-red fluorescent protein with fast maturation and reduced oligomerization tendency from *Entacmaea quadricolor* (Anthozoa, Actinaria). *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11646-51 (2002).

31. Petersen J, Wilmann PG, Beddoe T, Oakley AJ, Devenish RJ, Prescott M, Rossjohn J. The 2.0-Å crystal structure of eqFP611, a far red fluorescent protein from the sea anemone *Entacmaea quadricolor*. *J. Biol. Chem*. **278**, 44626-31 (2003).

32. Karasawa S, Araki T, Nagai T, Mizuno H, Miyawaki A. Cyan-emitting and orange-emitting fluorescent proteins as a donor/acceptor pair for fluorescence resonance energy transfer. *Biochem. J*. **381**, 307-312 (2004).

33. Kikuchi A, Fukumura E, Karasawa S, Mizuno H, Miyawaki A, Shiro Y. Structural characterization of a thiazoline-containing chromophore in an orange fluorescent protein, monomeric Kusabira Orange. *Biochemistry* **47**, 11573-80 (2008).

34. Merzlyak EM, Goedhart J, Shcherbo D, Bulina ME, Shcheglov AS, Fradkov AF, Gaintzeva A, Lukyanov KA, Lukyanov S, Theodorus W J Gadella TWJ, Chudakov DM. Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods* **4**, 555-7 (2007).

35. Tretyakova YA, Pakhomov AA, Martynov VI. Chromophore structure of the kindling fluorescent protein asFP595 from *Anemonia sulcata*. *J. Am. Chem. Soc*. **129**, 7748-9 (2007).

36. Ryoko Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12651-6 (2002).

37. Wiedenmann J, Ivanchenko S, Oswald F, Schmitt F, Röcker C, Salih A, Spindler K, Nienhaus GU. EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15905-10 (2004).

38. Hayashi I, Mizuno H, Tong KI, Furuta T, Tanaka F, Yoshimura M, Miyawaki A, Ikura M. Crystallographic evidence for water-assisted photo-induced peptide cleavage in the stony coral fluorescent protein Kaede. *J. Mol. Biol*. **372**, 918-26 (2007).

39. Gurskaya NG, Verkhusha VV, Shcheglov AS, Staroverov DB, Chepurnykh TV, Fradkov AF, Lukyanov S, Lukyanov KA. Engineering of a monomeric green-to-red photoactivatable fluorescent protein induced by blue light. *Nat. Biotechnol*. **24**, 461-4 (2006).

40. Tsutsui H, Karasawa S, Shimizu H, Nukina N, Miyawaki A. Semi-rational engineering of a coral fluorescent protein into an efficient highlighter. *EMBO rep*. **6**, 233-8 (2005).

41. Stiel AC, Trowitzsch S, Weber G, Martin Andresen M, Eggeling C, Hell SW, Jakobs S, Wahl MC. 1.8Å bright-state structure of the reversibly switchable fluorescent protein Dronpa guides the generation of fast switching variants. *Biochem. J*. **402**, 35-42 (2006).

42. Day RN & Davidson MW. The fluorescent protein revolution. 76-99 (2014).

43. Baird GS, Zacharias DA, Tsien RY. Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11984-9 (2000).

44. Heim R, Cubitt AB, Tsien RY. Improved green fluorescence. *Nature* **373**, 663-4 (1995).

45. Cormack BP, Valdivia RH, Falkow S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*. **173**, 33-8 (1996).

46. Ormo M, Cubitt AB, Kallio K, Gross LA, Tsien RY, Remington SJ. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science*. **273**, 1392-5 (1996).

47. Heim R, Prasher DC, Tsien RY. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12501-4 (1994).

48. Cubitt AB, Woollenweber LA, Heim R. Understanding structure-function relationships in the *Aequorea Victoria* green fluorescent protein. *Methods cell biol*. **58**, 19-30 (1999).

49. Heim R, Tsien RY. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol*. **6**, 178–82 (1996).

50. Bevis BJ, Glick BS. Rapidly maturing variants of the *Discosoma* red fluorescent protein (DsRed). *Nat. Biotechnol*. **20**, 83-7 (2002).

51. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY. A monomeric red fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7877-82 (2002).

52. Shaner NC, Campell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat. Biotechnol*. **22**, 1567-72 (2004).

53. Shaner NC, Lin MZ, McKeown MR, Steinbach PA, Hazelwood KL, Michael W Davidson MW, Tsien RY. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods*. **5**, 545-51 (2008).

54. Subach OM, Gundorov IS, Yoshimura M, Subach FV, Jinghang Zhang J, Grüenwald D, Souslova EA, Chudakov DM, Verkhusha VV. Conversion of red fluorescent protein into a bright blue probe. *Chem. Biol*. **15**, 1116-24 (2008).

55. Subach OM, Malashkevich VN, Zencheck WD, Morozova KS, Piatkevich KD, Almo SC, Verkhusha VV. Structural characterization of acylimine-containing blue and red chromophores in mTagBFP and TagRFP fluorescent proteins. *Chem. Biol*. **17**, 333–41 (2010).

56. Shcherbo D, Merzlyak EM, Chepurnykh TV, Fradkov AF, Ermakova GV, Solovieva EA, Lukyanov KA, Bogdanova EA, Zaraisky AG, Lukyanov S, Chudakov DM. Bright far-red fluorescent protein for whole-body imaging. *Nat. Methods*. **4**, 741-6 (2007).

57. Lin MZ, McKeown MR, Ng H, Aguilera TA, Shaner NC, Campbell RE, Adams SR, Gross LA, Ma W, Alber T, Tsien RY. Autofluorescent proteins with excitation in the optical window for intravital imaging in mammals. *Chem. Biol*. **16**, 1169–79 (2009).

58. Ibraheem A, Campbell RE. Designs and applications of fluorescent protein-based biosensors. *Curr. Opin. Chem. Biol*. **14**, 30–6 (2010).

59. Müller-Taubenberger A, Anderson KI. Recent advances using green and red fluorescent protein variants. *Appl. Microbiol. Biotechnol*. **77**, 1-12 (2007).

60. Lam A, St-Pierre1 F, Gong Y, Marshall JD, Cranfill PJ, Baird MA, McKeown MR, Wiedenmann J, Davidson MW, Schnitzer MJ, Tsien RY, Lin MZ. Improving FRET dynamic range with bright green and red uorescent proteins. *Nat. Methods*. **9**, 1005–12 (2012).

61. Bajar BT, Wang ES, Zhang S, Lin MZ, Chu J. A guide to fluorescent protein FRET pairs. *Sensors* **16**, 1488 (2016).

62. Bajar BT, Wang ES, Lam AJ, Kim BB, Jacobs CL, Howe ES, Davidson MW, Lin MZ, Chu J. Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting. *Sci. Rep*. **6**, 20889 (2016).

63. Chen Z, Ren W, Wright QE, Ai H. Genetically encoded fluorescent probe for the selective detection of peroxynitrite. *J. Am. Chem. Soc*. **135**, 14940-3 (2013).

64. Kredel S, Oswald F, Nienhaus K, Deuschle K, Röcker C, Wolff M, Heilker R, Nienhaus GU, Wiedenmann J. mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures. *PLoS ONE*. **4**, e4391 (2009).

65. Crowe J, Dobeli H, Gentz R, Hochuli E, Stiiber D, Henco K. 6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification. *Methods Mol. Biol.* **31**, 371-87 (1994).

66. Bergfors TM. Protein Crystallization. 97-114 (2009).

67. Kabsch, W. XDS. *Acta Crystallogr D Struct Biol.* **66**, 125-32 (2010).

68. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-74 (2007).

69. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Struct Biol.* 60, 2126-32 (2004).

70. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Struct Biol.* **67**, 235-42 (2011).

71. Tanaka S, Ataka M. Protein crystallization induced by polyethylene glycol: A model study using apoferritin. *J Chem Phys*. **117**, 3504-10 (2002).

72. Akerboom J, Calderón NC, Tian L, Wabnig S, Prigge M, Tolö J, Gordus A, Orger MB, Severi KE, Macklin JJ, Patel R, Pulver SR, Wardill1 TJ, Fischer E, Schüler C, Chen T, Sarkisya KS, Marvin JS, Bargmann CI, Kim DS, Kügler S, Lagnado L, Hegemann P, Gottschalk A, Eric R. Schreiter ER, Looger LL. Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front. Mol. Neurosci*. **6**, 1-29 (2013).

73. Zhao Y, Araki S, Wu J, Teramoto T, Chang Y, Nakano M, Abdelfattah AS, Fujiwara M, Ishihara T, Nagai T, Campbell RE. An expanded palette of genetically encoded Ca^{2+} indicators. *Science*. **333**, 1888-91 (2011).

74. Wongthai P, Hagiwara K, Miyoshi Y, Wiriyasermkul P, Wei L, Ohgaki R, Kato I, Hamase K, Nagamori S, Kanai Y. Boronophenylalanine, a boron delivery agent for boron neutron capture therapy, is transported by $ATB^{0,+}$, LAT1 and LAT2. *Cancer Sci.* **106**, 279–286 (2015).

75. Bergfors T. Seeds to crystals. *J. Struct. Biol*. **142**, 66–76 (2013).

76. Kowalinski E, Bange G, Wild K, Sinning I. Expression, purification, crystallization and preliminary crystallographic analysis of the proliferation-associated protein Ebp1. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63**, 768-70 (2007).

77. Hall DG. Boronic Acids. 1-100 (2005).

78. Wang F, Niu W, Guo J, Schultz PG. Unnatural amino acid mutagenesis of fluorescent proteins. *Angew. Chem. Int. Edit*. **51**, 10132-5 (2012).

79. Chen Z, Tian Z, Kallio K, Oleson AL, Ji A, Borchardt D, Jiang D, Remington SJ, Ai H. The N−B interaction through a water bridge: understanding the chemoselectivity of a fluorescent protein based probe for peroxynitrite. *J. Am. Chem. Soc.* **138**, 4900-7 (2016).

80. Djinovic-Carugo K, Carugo O. Missing strings of residues in protein crystal structures. *Intrinsically Disord. Proteins* **3**, e1095697 (2015).